

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 16:25:00 ; Search time 45 Seconds
(without alignments)
3814.264 Million cell updates/sec

Title: US-10-658-989A-4

Perfect score: 3070

Sequence: 1 GSGPGVGRGPGPGPAGA.....PPSGDAGPPGPGPAGKEG 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1461	4	O76045
2	2966	95.6	1464	4	Q8N473
3	2872	93.6	1453	11	Q63079
4	2857	93.1	1453	11	Q81009
5	2528	82.3	1450	13	Q9YIB4
6	2525	82.2	1445	13	Q93251
7	2525	82.2	1449	13	Q802B5
8	2484.5	80.9	1447	13	Q9IB91
9	2231	72.7	1449	13	Q910C0
10	2225	72.5	1487	6	O77753
11	2224	72.4	1160	4	Q14046
12	2224	72.4	1487	4	Q14047
13	2220	72.3	1418	6	Q28396
14	2195	71.5	1269	13	Q7C227
15	2192	71.4	1419	11	Q63123
16	2192	71.4	1419	11	Q80X38

17	2192	71.4	1419	11	Q80VY3
18	2192	71.4	1442	11	O62031
19	2192	71.4	1442	11	O62033
20	2192	71.4	1459	11	O62032
21	2185	71.2	1420	13	Q90X37
22	2147	69.9	1486	13	Q91717
23	2145	69.9	1486	13	Q7ZII6
24	2140	69.7	1418	13	Q9W7R9
25	2128	69.3	1491	13	Q91718
26	2126	69.3	1491	13	Q7ZTM4
27	2105	68.6	1458	13	Q910B9
28	1953	63.6	1347	4	Q96Q83
29	1923	62.6	1497	11	Q7TMS0
30	1915	62.4	1314	11	Q8CGA5
31	1910	62.2	1497	11	Q61431
32	1899	61.9	998	11	O8CFM4
33	1899	61.9	1222	11	Q8KI73
34	1899	61.9	1464	11	Q8BLM4
35	1899	61.9	1464	11	Q7TJ32
36	1892.5	61.6	1464	11	Q8BKV2
37	1886	61.4	1366	4	Q7Z586
38	1885	61.4	1366	4	Q15177
39	1875	61.1	1352	13	Q90Y70
40	1823	59.4	1163	4	Q8N6U4
41	1798	58.6	1346	13	Q8UUG3
42	1798	58.6	1352	13	Q8UUG4
43	1792.5	58.4	1346	13	Q801M5
44	1635	53.3	1258	13	Q8AW11
45	1550.5	50.5	1414	5	Q26634

ALIGNMENTS

RESULT 1

O76045 PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pro alpha 1(I) collagen.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88399734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;

Q80VY3 mus musculus
Q62031 mus musculus
Q62033 mus musculus
Q62032 mus musculus
Q90W37 gallus galli
Q91717 xenopus lae
Q7ZT16 xenopus lae
Q9W7R9 cynops pyrr
Q91718 xenopus lae
Q7ZTM4 xenopus lae
Q910B9 oncorhynch
Q96Q83 homo sapien
Q7TMS0 mus musculus
Q8CGA5 mus musculus
Q61431 mus musculus
Q8CFM4 mus musculus
Q8KI73 mus musculus
Q8BLM4 mus musculus
Q7TJ32 mus musculus
Q8BKV2 mus musculus
Q7Z586 homo sapien
Q15177 homo sapien
Q90Y70 brachydanio
Q8N6U4 homo sapien
Q8UUG3 oncorhynch
Q8UUG4 oncorhynch
Q801M5 xenopus lae
Q8AW11 brachydanio
Q26634 strongyloce

"Highly conserved sequences in the 3'-untranslated region of the COL1A1 gene bind cell-specific nuclear proteins.;"
 FEBS Lett. 279:9-13(1991).
 [5]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=92157916; PubMed=1787829;
 RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
 RA Olsen A.S., Prockop D.J.;
 RA "Completion of the last half of the structure of the human gene for
 RT the pro alpha 1(I) chain of type I procollagen (COL1A1).";
 RL Matrix 11:375-379(1991).
 RN [6]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98107942; PubMed=9443882;
 RA Korkko J., Ala-Kokko L., De Paepe A., Nuytinck L., Earley J.,
 RA Prockop D.J.;
 RA "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
 RT scanning by conformation-sensitive gel electrophoresis identifies only
 RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
 RT identification of common sequences of null-allele mutations.;"
 RL Am. J. Hum. Genet. 62:98-110(1998).
 RN [7]
 RN SEQUENCE FROM N.A.
 RP Korkko J.M., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
 RA Ala-Kokko L.;
 RA Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF017178; AAB94054.2; -
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR002181; Fibrinogen C.
 DR InterPro; IPR000885; Fib collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 2.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS0184; VWF_C_2; 1.
 DR Collagen.
 KW SEQUENCE 1461 AA; 138630 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 96.6%; Score 2966; DB 4; Length 1461;
 Best Local Similarity 95.6%; Pred. No. 1.3e-169;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GSEGPVGRGEPGPPGAGAGPAGDPGADGEPGAKGADGAPGAGPFGAGPSPGE 60
 DB 359 GSEGPVGRGEPGPPGAGAGPAGDPGADGEPGAKGADGAPGAGPFGAGPSPGE 418
 QY 61 GGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVPVGGPPGAGGKPGAGGPGTGLP 120
 DB 419 GGGPPGPKGNSGEPGAPGSKGDTGAKGEPGVPVGGPPGAGGKPGAGGPGTGLP 478
 QY 121 GPPGEGGPGSRGFPAGDGVAGPKGFPAGERSGFPAGPKGSPGAEAGLPGAKGLT 180
 DB 479 GPPGEGGPGSRGFPAGDGVAGPKGFPAGERSGFPAGPKGSPGAEAGLPGAKGLT 538
 QY 181 GPGSPGPPGKGTGPPGAGEDRGPDPGPPGARGAGVWGPPGKAGERPAGGVP 240
 DB 539 GPGSPGPPGKGTGPPGAGQDGRGPPGPPGARGAGVWGPPGKAGERPAGGVP 598
 QY 241 GPPGAVGPAKDGAEAGPAGPAGPAGERSGFPAGSGPAGPAGPAGPAGPAGP 300
 DB 599 GPPGAVGPAKDGAEAGPAGPAGPAGERSGFPAGSGPAGPAGPAGPAGPAGP 658
 QY 301 GVPDGLGAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
 DB 659 GVPDGLGAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 718

QY 361 GAPGLEMPGERGAAGLPGPKGDRDAGPKGADGSGPKDGVRLGTGPIGPPGAGAPGDK 420
 DB 719 GAPGLQMPGERGAAGLPGPKGDRDAGPKGADGSGPKDGVRLGTGPIGPPGAGAPGDK 778
 QY 421 GSEGPSPAGPTGARGAPGDRGEPGPPGAGPAGPAGDGEPCAKGECDAKGDAGPP 480
 DB 779 GSEGPSPAGPTGARGAPGDRGEPGPPGAGPAGPAGDGEPCAKGECDAKGDAGPP 838
 QY 481 GPAGPAGPPGPTGVDVGAOPGAKGARGSPGATGTFGAAGRVGPPGPGSDAGPPGPPA 540
 DB 839 GPAGPAGPPGPTGVDVGAOPGAKGARGSPGATGTFGAAGRVGPPGPGSDAGPPGPPA 598
 QY 541 GKEG 544
 DB 899 GKEG 902
 RESULT 2
 Q8N473 PRELIMINARY; PRT; 1464 AA.
 ID Q8N473
 AC Q8N473; 2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RC Tissue=Brain;
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC036531; AH36531.1; -
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR002181; Fibrinogen C.
 DR InterPro; IPR000885; Fib collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 2.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS0184; VWF_C_2; 1.
 DR Hypothetical protein; Collagen.
 KW SEQUENCE 1464 AA; 139011 MW; B0581F8D1C89DDE8 CRC64;
 Query Match 96.6%; Score 2966; DB 4; Length 1464;
 Best Local Similarity 95.6%; Pred. No. 1.3e-169;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GSEGPVGRGEPGPPGAGAGPAGDPGADGEPGAKGADGAPGAGPFGAGPSPGE 60
 DB 362 GSEGPVGRGEPGPPGAGAGPAGDPGADGEPGAKGADGAPGAGPFGAGPSPGE 421
 QY 61 GGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVPVGGPPGAGGKPGAGGPGTGLP 120
 DB 422 GPGGPPGPKNSGEPGAPGSKGDTGAKGEPGVPVGGPPGAGGKPGAGGPGTGLP 481
 QY 121 GPPGEGGPGSRGFPAGDGVAGPKGFPAGERSGFPAGPKGSPGAEAGLPGAKGLT 180
 DB 482 GPPGEGGPGSRGFPAGDGVAGPKGFPAGERSGFPAGPKGSPGAEAGLPGAKGLT 541
 QY 181 GPGSPGPPGKGTGPPGAGEDRGPDPGPPGARGAGVWGPPGKAGERPAGGVP 240
 DB 542 GPGSPGPPGKGTGPPGAGQDGRGPPGPPGARGAGVWGPPGKAGERPAGGVP 601

QY	241	GPFDVAGVAGKDDGEAGAGAGPPGPGAPGAPGAGEEGEPAGSPGFEGLPAGPAGPPGEGAGKPGGE	300			
Db	602	GPFDVAGVAGKDDGEAGAGAGPPGPGAPGAPGAGEEGEPAGSPGFEGLPAGPAGPPGEGAGKPGGE	661			
QY	301	GVPGDLGAPGSGAGGEPGPGGEGVGPAGPCGADGAPGDDGAKGDAGAPGAPGSE	360			
Db	662	GVPGDLGAPGSGAGGEPGPGGEGVGPAGPCGADGAPGDDGAKGDAGAPGAPGSE	721			
QY	361	GAPGLEGMPPGERGAAGLPKPKXDRGDAGPKGADGSPCKDVGRLGTGPIGPPGAPAGPDK	420			
Db	722	GAPGLEGMPPGERGAAGLPKPKXDRGDAGPKGADGSPCKDVGRLGTGPIGPPGAPAGPDK	781			
QY	421	GESGSGPAGPTGARGAGCDGRGEPGPGAPGAGPCGADGEPGAKGCEPGDAGAKGDAGP	480			
Db	782	GESGSGPAGPTGARGAGCDGRGEPGPGAPGAGPCGADGEPGAKGCEPGDAGAKGDAGP	841			
QY	481	GPAGPAGPPGPTGVDVGAFCAGKAGSAGGATGPPCAAGRVGPPGSPGADGPPGPPGPA	540			
Db	842	GPAGPAGPPGPTGVDVGAFCAGKAGSAGGATGPPCAAGRVGPPGSPGADGPPGPPGPA	901			
QY	541	GKEG 544				
Db	902	GKEG 905				
RESULT 3						
Q63079 PRELIMINARY; PRT; 1453 AA.						
AC	Q63079	PRELIMINARY; PRT; 1453 AA.				
DT	01-JUN-1996	(T=EMBLrel. 01, Created)				
DT	01-JUN-1998	(T=EMBLrel. 06, Last sequence update)				
DT	01-OCT-2003	(T=EMBLrel. 25, Last annotation update)				
DE	Collagen alpha1 (Fragment).					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
ON	NCBI_TaxID=10116;					
RX	[1]					
RP	SEQUENCE OF 1-1092 FROM N.A.					
RA	STRAIN=Sprague-Dawley; TISSUE=Tooth;					
RA	Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;					
RT	"Expression of Collagen alpha1(I) mRNA variants during Tooth and Bone					
RT	Formation in the Rat.;"					
RL	J. Dent. Res. 0:0-0(0).					
RL	[2]					
RP	SEQUENCE FROM N.A.					
RA	STRAIN=Sprague-Dawley; TISSUE=Tooth;					
RA	Wurtz T.;					
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; Z78279; CAB01633.1; -.					
DR	GO; GO:0005581; C:collagen; IEA.					
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.					
DR	InterPro; IPR008161; Clg_helix.					
DR	InterPro; IPR008160; Collagen.					
DR	InterPro; IPR002181; Fibrinogen_C.					
DR	InterPro; IPR000885; Fib_collagen_C.					
DR	InterPro; IPR001007; VWF_C.					
DR	Pfam; PF01410; COLFI; 1.					
DR	Pfam; PF01391; Collagen; 18.					
DR	ProDom; ED000007; Clg_helix; 3.					
DR	ProDom; PD002078; Fib_collagen_C; 1.					
DR	SMART; SM00038; COLFI; 1.					
DR	SMART; SM00214; VWC; 1.					
DR	PROSITE; PS01208; VWF_C; 1.					
DR	PROSITE; PS01208; VWF_C; 1.					
DR	Collagen.					
KW	NON_TER					
FT	NON_TER					
SQ	SEQUENCE 1453 AA; 137887 MW; E6896BDC19A4A1D8 CRC64;					
Query Match 93.6%; Score 2872; DB 11; Length 1453;						
Best Local Similarity 92.1%; Pred.No.5,2e-164;						
Matches 501; Conservative 26; Mismatches 17; Indels 0; Gaps						

Qy	1	GSGPGEVGRGSEPPPGAGNAAGPAGDCEPGAAGADCAFGTAGAPGGPGARGPSGP	60
Dd	351	GSGEPGVGRGSEPPPGAGNAAGPAGDCEPGAAGADCAFGTAGAPGGPGARGPSGP	410
Qy	61	GPGGPPPKGDSBEPKAPSKGDTGAKEGPEPVGVVEPPGPAEGEKGARGEGPGTGLP	120
Dd	411	GPSGAPCPKTSBEPKAPSKGDTGAKEGPEPVGVVEPPGPAEGEKGARGEGPGSGLP	470
Qy	121	GPGERGGPSRPFPGADGVAGPKPGPAGERGSFQFPACPKSPGHAGRPGEAGLPGAAGLT	180
Dd	471	GPGERGGPSRPFPGADGVAGPKPGPAGERGSFQFPACPKSPGHAGRPGEAGLPGAAGLT	530
Qy	181	GSPGSDPDGKTGPPGPGAGEDRPPGPPPCARGEAGVMGFPPGPKAAGEPGKAGERGVP	240
Dd	531	GSPGSDPDGKTGPPGPGAGEDRPPGPPPCARGEAGVMGFPPGPKAAGEPGKAGERGVP	590
Qy	241	GPPGAVGPAKDGAEAGAEGPPGPAAGERGBEGPAGSPQFEGLUPGAGPPPGEAGKPGEE	300
Dd	591	GPPGAVGPAKDGAEAGAEGPPGPAAGERGBEGPAGSPQFEGLUPGAGPPPGEAGKPGEE	650
Qy	301	GVPGDLGARPSGNARBEPPGPRGIVTGPPGPAAGPAGDAGCAPGDGAKGADAGAPGASG	360
Dd	651	GVPGDLGARPSGNARBEPPGPRGIVTGPPGPAAGPAGDAGCAPGDGAKGADAGAPGASG	710
Qy	361	GAPLEGMPGERGAAGLPGPKDRGDAGPKGADGSPKODGVYRGITGPIPPGPAAGPDK	420
Dd	711	GAPLEGMPGERGAAGLPGPKDRGDAGPKGADGSPKODGVYRGITGPIPPGPAAGPDK	770
Qy	421	GESGFSGPAPTGARGAPDRBERGPPGPPGPAAGPAGDAGBPAGKBPBGDAKAGDAGPP	480
Dd	771	GESGFSGPAPTGARGAPDRBERGPPGPPGPAAGPAGDAGBPAGKBPBGDAKAGDAGPP	830
Qy	481	GPAGPAGPPGPIGVDPGAPKAGKARGSGAGPPCATCFPGAAGRVGPPGSGDAGPPGPPGA	540
Dd	831	GPAGPAGPPGPIGVDPGAPKAGKARGSGAGPPCATCFPGAAGRVGPPGSGDAGPPGPPGA	890
Qy	541	GKEG 544	
Dd	891	GKEG 894	
 RESULT 4 Q810J9 PRELIMINARY; PRT; 1453 AA.			
AC	Q810J9		
DT	01-JUN-2003	(TrEMBLrel. 24, Created)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N; TISSUE=Colon;		
RA	Strausberg R.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC050014; AAH50014.1; -		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	InterPro; IPR008160; Collagen.		
DR	InterPro; IPR002181; Fibrinogen_C.		
DR	InterPro; IPR000885; Fib_collagen_C.		
DR	InterPro; IPR01007; VWF_C.		
DR	Pfam; PF01410; COLFI; 1.		
DR	Pfam; PF01391; Collagen; 18.		
DR	ProDom; PD002078; Fib_collagen_C; 1.		
DR	SMART; SM00036; COLFI; 1.		
DR	SMART; SM00214; VWC; 1.		
DR	PROSITE; PS01208; VWF_C_1; 1.		
DR	PROSITE; PS01184; VWF_C_2; 1.		
KW	Hyvothermal protein.		

```

SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BB9A1D5EA CRC64;
Query Match 93.1%; Score 2857; DB 11; Length 1453;
Best Local Similarity 91.4%; Pred. No. 4.1e-163;
Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;

QY 1 GSGGPEGVGRRGPPGPPAGAGAGAGDAGDGEFGAKGADGAPGAGPFGAGRGSGPE 60
Db 351 GSEGPQGVRRGPPGPPAGAGAGAGDAGDGEFGAKGADGAPGAGPFGAGRGSGPQ 410
QY 61 GPGGPPGPKDSGPPGAPGSKGDTGAKGEPGVGVGPPGAGGKPGAGRGPGTGLP 120
Db 411 GPGGPPGPKDSGPPGAPGSKGDTGAKGEPGVGVGPPGAGGKPGAGRGPGTGLP 120
QY 121 GPPGRRGPPGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 180
Db 471 GPPGRRGPPGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 180
QY 181 GSPGSPGDPGKTGPPGAGDGRGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGP 240
Db 531 GSPGSPGDPGKTGPPGAGDGRGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGP 240
QY 241 GPPGAVGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 300
Db 591 GPPGAVGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 300
QY 301 GVPGLGAPGSPGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 360
Db 651 GVPGLGAPGSPGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 360
QY 361 GAPGLEGMPPGRRGAGLPGPKGDRGADGPKGADGSPGKGVRLGTGPPGAGRGSG 420
Db 711 GAPGLEGMPPGRRGAGLPGPKGDRGADGPKGADGSPGKGVRLGTGPPGAGRGSG 420
QY 421 GESGSPGAGTGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 480
Db 771 GESGSPGAGTGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 480
QY 481 GPAGPAGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 540
Db 831 GPAGPAGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 540
QY 541 GKEG 544
Db 891 GKEG 894

RESULT 5
QY1B4 PRELIMINARY; PRT; 1450 AA.
AC QY1B4
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RA Asahina K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.

```

```

DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWF; 1.
DR PROSITE; PS01208; VWF_1; 1.
DR PROSITE; PS01184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137564 MW; ABF8A74841B87B7C CRC64;
Query Match 82.3%; Score 2528; DB 13; Length 1450;
Best Local Similarity 80.3%; Pred. No. 1.8e-143;
Matches 437; Conservative 43; Mismatches 64; Indels 0; Gaps 0;

QY 1 GSEGEVGVRRGPPGPPAGAGAGDAGDGEFGAKGADGAPGAGPFGAGRGSGPE 60
Db 348 GSEGEVGVRRGPPGPPAGAGAGDAGDGEFGAKGADGAPGAGPFGAGRGSGPQ 407
QY 61 GPGGPPGPKDSGPPGAPGSKGDTGAKGEPGVGVGPPGAGGKPGAGRGPGTGLP 120
Db 408 GPGGPPGPKDSGPPGAPGSKGDTGAKGEPGVGVGPPGAGGKPGAGRGPGTGLP 120
QY 121 GPPGRRGPPGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 180
Db 468 GPPGRRGPPGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 180
QY 181 GSPGSPGDPGKTGPPGAGDGRGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGP 240
Db 528 GSPGSPGDPGKTGPPGAGDGRGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGP 240
QY 241 GPPGAVGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 300
Db 588 GPPGAVGPPGAGDGVAGPKGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 300
QY 301 GVPGLGAPGSPGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 360
Db 648 GVPGLGAPGSPGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 360
QY 361 GAPGLEGMPPGRRGAGLPGPKGDRGADGPKGADGSPGKGVRLGTGPPGAGRGSG 420
Db 708 GAPGLEGMPPGRRGAGLPGPKGDRGADGPKGADGSPGKGVRLGTGPPGAGRGSG 420
QY 421 GESGSPGAGTGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 480
Db 768 GESGSPGAGTGARGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 480
QY 481 GPAGPAGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 540
Db 828 GPAGPAGPPGPPGAGRGSGPPGAGRGSGPPGAGRGSGPPGAGRGSGPGLP 540
QY 541 GKEG 544
Db 888 GKEG 891

RESULT 6
QY1B4 PRELIMINARY; PRT; 1445 AA.
AC QY1B4
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Alpha 1 type I collagen.
OS Rana catesbeiana (bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99294154; PubMed=10367734;
RA Asahina K., Uton R., Obara M., Yoshizato K.;

```


RESULT 8
Q9IB91 PRELIMINARY; PRT; 1447 AA.

AC Q9IB91
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Type I collagen alpha 1.
GN COL1A1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OX Xenopodinae; Xenopus.
NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RA "Expression and characterization of Xenopus type I collagen alpha 1
RT (COL1A1) during embryonic development.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB034701; BAA94972.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPRO08161; Clg_helix.
DR InterPro; IPRO08160; Collagen.
DR InterPro; IPRO00885; Fib_collagen_C.
DR InterPro; IPRO01007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PSS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1447 AA; 137446 MW; A4A6DD2B4158E38B CRC64;

Query Match 80.9%; Score 2484.5; DB 13; Length 1447;
Best Local Similarity 79.8%; Pred. No. 7.3e-141;
Matches 433; Conservative 50; Mismatches 60; Indels 1; Gaps 1;

Qy 1 GSSEGVGVEGPPGPAGAGPAGDPADGGFPGKAGDAPGIAGAFPCFGARPSGE 60
Db 348 GSDPGGRCEPCAPGAQAAGSPGSGDGQPGAAGATGATGAGIAGAAGFFGARGAPCAQ 407

Qy 61 GPGGPGPKDSERCAPSKGDTGAKGEPGVGVEGPPGAGEGKGCAGERGEPTGLP 120
Db 408 GPGSPPKXNNGEPGAQNKKGAGAKGFPAGVQGVPGPSGEKGRSSEP-PAGFP 466

Qy 121 GPPGERGGPSRFPAGADVAGPKGPAERGSPGPAKPSPCEAGRICEAGLPKAGLT 180
Db 467 GPAGERGGPSRRFPAGSDCAGSKGAPGEPVGPAGPKSSGESRPGECPLFCAGKLT 526

Qy 181 GPSGSPGDKTTPGPAGEDRDPPGPPGARGBAWGMFPFGKAAGEFKKAGERVVP 240
Db 527 GPSGSPGSDKTTPAGAQDGRAGPPGPPGARGQSVMGFPGKAAGEFKKKEGVA 586

Qy 241 GPGAVGPAGCKGEAAGEPPGAPGAPGERGEPAGSPCFGLPGPAGPGEACKPEE 300
Db 587 GPGAVGLPKDGDAGAQGPDPGAPGERGQQGAPGPFQLPGSPGPAGESKPGEQ 646

Qy 301 GVPDGLGAPGSPGARGEPPGGERGVGPPGAPGADGAPCDGAKGDAGAPGSE 360
Db 647 GVPDVGPSFAGARGERFPFGRGAQQGPPGPPQCARGSNCPNDGAKGEAGAAPGGQ 706

Qy 361 GAPGLEMPGERGNAGLPGPKDRGAPGKAGDGSFGDKVGLTGPIGPCGAPAGDK 420
Db 707 GSPCLQMPPERSSSLFGAKRGDQGVKSGDGTGKDGVRGLTGPIGPPGAPGDK 766

Qy 421 GESFPSGAPGFTGARGAPCDRGPFPFPAGFAGFPAGDFPCAGKGEPCDAGAKDAGFP 480

[illegible]

alpha 1(II) chain is more similar to the alpha 1(I) chain than two other alpha chains of fibrillar collagens.";
 RL Biochem. J. 262:521-528 (1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=89325561; PubMed=2753125;
 RA Viikula M., Peltonen L.;
 RT "Structural analyses of the polymorphic area in type II collagen gene.";
 RL FEBS Lett. 250:171-174 (1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=91118411; PubMed=2081599;
 RA Ryan M.C., Sieraaki M., Sandell L.J.;
 RT "The human type II procollagen gene: identification of an additional protein-coding domain and location of potential regulatory sequences in the promoter and first intron.";
 RL Genomics 8:41-48 (1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=91153296; PubMed=1999183;
 RA Huang M.C., Seyer J.M., Thompson J.P., Spinella D.G., Cheah K.S., Kang A.H.;
 RT "Genomic organization of the human procollagen alpha 1(II) collagen gene.";
 RL Eur. J. Biochem. 195:593-600 (1991).
 RN [6]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=92344585;
 RA Viikula M., Metsaranta M., Syr nen A.C., Ala-Kokko L., Vuorio E., Peltonen L.;
 RT "Structural analysis of the regulatory elements of the type-II procollagen gene. Conservation of promoter and first intron sequences between human and mouse.";
 RL Biochem. J. 285:0-0 (0).
 RN [7]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=97104294; PubMed=8948452;
 RA Ala-Kokko L., Kvist A.P., Metsaranta M., Kivirikko K.I., "Conservation of the sizes of 53 introns and over 100 intronic sequences for the binding of common transcription factors in the human and mouse genes for type II procollagen (COL2A1).";
 RL and mouse genes for type II procollagen (COL2A1).";
 RN [7]

Query Match 72.4%; Score 2224; DB 4; Length 1487;
Best Local Similarity 72.6%; Pred. No. 2.8e-125;
Matches 395; Conservative 38; Mismatches 111; Indels 0; Gaps 0;
QY 1 GSEGEVGRGEPGPAGAGAGDPGDGDFPGAKGADGAPGTAGAPFGARGPSGPE 60

[illegible]

RESULT 15

```

000001 15
000002
000003
000004
000005
000006
000007
000008
000009
000010
000011
000012
000013
000014
000015
000016
000017
000018
000019
000020
000021
000022
000023
000024
000025
000026
000027
000028
000029
000030
000031
000032
000033
000034
000035
000036
000037
000038
000039
000040
000041
000042
000043
000044
000045
000046
000047
000048
000049
000050
000051
000052
000053
000054
000055
000056
000057
000058
000059
000060
000061
000062
000063
000064
000065
000066
000067
000068
000069
000070
000071
000072
000073
000074
000075
000076
000077
000078
000079
000080
000081
000082
000083
000084
000085
000086
000087
000088
000089
000090
000091
000092
000093
000094
000095
000096
000097
000098
000099
000100
000101
000102
000103
000104
000105
000106
000107
000108
000109
000110
000111
000112
000113
000114
000115
000116
000117
000118
000119
000120
000121
000122
000123
000124
000125
000126
000127
000128
000129
000130
000131
000132
000133
000134
000135
000136
000137
000138
000139
000140
000141
000142
000143
000144
000145
000146
000147
000148
000149
000150
000151
000152
000153
000154
000155
000156
000157
000158
000159
000160
000161
000162
000163
000164
000165
000166
000167
000168
000169
000170
000171
000172
000173
000174
000175
000176
000177
000178
000179
000180
000181
000182
000183
000184
000185
000186
000187
000188
000189
000190
000191
000192
000193
000194
000195
000196
000197
000198
000199
000200
000201
000202
000203
000204
000205
000206
000207
000208
000209
000210
000211
000212
000213
000214
000215
000216
000217
000218
000219
000220
000221
000222
000223
000224
000225
000226
000227
000228
000229
000230
000231
000232
000233
000234
000235
000236
000237
000238
000239
000240
000241
000242
000243
000244
000245
000246
000247
000248
000249
000250
000251
000252
000253
000254
000255
000256
000257
000258
000259
000260
000261
000262
000263
000264
000265
000266
000267
000268
000269
000270
000271
000272
000273
000274
000275
000276
000277
000278
000279
000280
000281
000282
000283
000284
000285
000286
000287
000288
000289
000290
000291
000292
000293
000294
000295
000296
000297
000298
000299
000300
000301
000302
000303
000304
000305
000306
000307
000308
000309
000310
000311
000312
000313
000314
000315
000316
000317
000318
000319
000320
000321
000322
000323
000324
000325
000326
000327
000328
000329
000330
000331
000332
000333
000334
000335
000336
000337
000338
000339
000340
000341
000342
000343
000344
000345
000346
000347
000348
000349
000350
000351
000352
000353
000354
000355
000356
000357
000358
000359
000360
000361
000362
000363
000364
000365
000366
000367
000368
000369
000370
000371
000372
000373
000374
000375
000376
000377
000378
000379
000380
000381
000382
000383
000384
000385
000386
000387
000388
000389
000390
000391
000392
000393
000394
000395
000396
000397
000398
000399
000400
000401
000402
000403
000404
000405
000406
000407
000408
000409
000410
000411
000412
000413
000414
000415
000416
000417
000418
000419
000420
000421
000422
000423
000424
000425
000426
000427
000428
000429
000430
000431
000432
000433
000434
000435
000436
000437
000438
000439
000440
000441
000442
000443
000444
000445
000446
000447
000448
000449
000450
000451
000452
000453
000454
000455
000456
000457
000458
000459
000460
000461
000462
000463
000464
000465
000466
000467
000468
000469
000470
000471
000472
000473
000474
000475
000476
000477
000478
000479
000480
000481
000482
000483
000484
000485
000486
000487
000488
000489
000490
000491
000492
000493
000494
000495
000496
000497
000498
000499
000500
000501
000502
000503
000504
000505
000506
000507
000508
000509
000510
000511
000512
000513
000514
000515
000516
000517
000518
000519
000520
000521
000522
000523
000524
000525
000526
000527
000528
000529
000530
000531
000532
000533
000534
000535
000536
000537
000538
000539
000540
000541
000542
000543
000544
000545
000546
000547
000548
000549
000550
000551
000552
000553
000554
000555
000556
000557
000558
000559
000560
000561
000562
000563
000564
000565
000566
000567
000568
000569
000570
000571
000572
000573
000574
000575
000576
000577
000578
000579
000580
000581
000582
000583
000584
00
```

Search completed: May 7, 2004, 16:29:49
Job time : 48 secs

[illegible]

Search completed: May 7, 2004, 16:29:49
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: May 7, 2004, 16:22:49 ; Search time 59 seconds
(without alignments)
2605.184 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGPGRVGEPPGPPGAGA.....PGPSGDAGPPGPPGAGREG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseqp_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2966	96.6	1057	3	AAy84541	AAy84541 Amino aci
2	2966	96.6	1057	3	AAy84544	AAy84544 A human c
3	2966	96.6	1058	3	AAy84403	AAy84403 Amino aci
4	2966	96.6	1107	2	AAr89472	AAr89472 Collagen/
5	2966	96.6	1107	3	AAy84540	AAy84540 Amino aci
6	2966	96.6	1161	7	ADe87050	ADe87050 Human pan
7	2966	96.6	1169	2	AAr89469	AAr89469 Collagen/
8	2966	96.6	1169	3	AAy84537	AAy84537 Amino aci
9	2966	96.6	1171	2	AAr89470	AAr89470 Collagen/
10	2966	96.6	1171	3	AAy84538	AAy84538 A chimeri
11	2966	96.6	1211	7	ADe87057	ADe87057 Human pan
12	2966	96.6	1226	7	ADe87062	ADe87062 Human pan
13	2966	96.6	1388	2	AAr89471	AAr89471 Collagen/
14	2966	96.6	1411	3	AAy56800	AAy56800 Human pre
15	2966	96.6	1461	5	ABg93947	ABg93947 Human pol
16	2966	96.6	1464	2	AAw68485	AAw68485 Human rec
17	2966	96.6	1464	4	AAw68485	AAw68485 Human pro
18	2966	96.6	1464	4	AAU14136	AAU14136 Human nov
19	2966	96.6	1464	5	ABH90764	ABH90764 Human tum
20	2966	96.6	1464	5	ABP68610	ABP68610 Human pan
21	2966	96.6	1464	6	ABU54471	ABU54471 Human tum
22	2966	96.6	1464	6	ABR47417	ABR47417 Breast ca
23	2966	96.6	1464	6	ABR92064	ABR92064 Human cer
24	2966	96.6	1464	7	ADD14142	ADD14142 Human src
25	2966	96.6	1464	7	ADD45059	ADD45059 Human Pro

26	2966	96.6	1464	7	ADD45055	ADD45055 Human Pro
27	2966	96.6	1464	7	ADD45051	ADD45051 Human Pro
28	2966	96.6	1464	7	ADe87048	ADe87048 Human pan
29	2958	96.4	1388	3	AAy84539	AAy84539 Amino aci
30	2936	95.6	1463	4	AAE02532	AAE02532 Bovine al
31	2920	95.1	1536	7	ADe87051	ADe87051 Human pan
32	2872	93.6	1453	7	ADD45053	ADD45053 Rat Prote
33	2872	93.6	1453	7	ADD45057	ADD45057 Rat Prote
34	2872	93.6	1453	7	ADD48341	ADD48341 Rat Prote
35	2872	93.6	1453	7	ADD45049	ADD45049 Rat Prote
36	2872	93.6	1453	7	ADD48337	ADD48337 Rat Prote
37	2872	93.6	1453	7	ADD48345	ADD48345 Rat Prote
38	2857	93.1	822	2	AAy06240	AAy06240 Mouse rec
39	2857	93.1	1453	5	ABg93948	ABg93948 Mouse pol
40	2816.5	91.7	1449	4	AAE02535	AAE02535 Porcine a
41	2791.5	90.9	1341	2	AAr71701	AAr71701 Collagen
42	2791.5	90.9	1341	3	AAy96122	AAy96122 Collagen
43	2791.5	90.9	1341	5	AAE16475	AAE16475 Human col
44	2791.5	90.9	1341	5	ABE80733	ABE80733 Collagen
45	2791.5	90.9	1341	5	ABB09625	ABB09625 Amino aci

ALIGNMENTS

RESULT 1
AAy84541
ID AAy84541 standard; protein; 1057 AA.
XX
AC AAy84541;
XX
DT 25-JUL-2000 (first entry)
XX
DE Amino acid sequence of a human collagen 1 (alpha1) protein.
XX
KW Extracellular matrix protein; self aggregation; hydroxylated proline;
trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS Homo sapiens.
XX
PN EP992586-A2.
XX
PD 12-APR-2000.
XX
PF 07-OCT-1999; 99EP-00119184.
XX
PR 09-OCT-1998; 98US-00169768.
XX
PA (USSU) US SURGICAL CORP.
XX
PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR WPI; 2000-259138/23.
XX
N-PSDB; AAA12502.
XX
PT Production of extracellular matrix proteins containing 4-trans-
hydroxyproline results in native self aggregating proteins, useful on
medical implants.
XX
PS Disclosure; Fig 27A-E; 260pp; English.
XX
CC The specification describes a method for producing an extracellular
matrix protein or its fragment. The extracellular matrix protein is
capable of self aggregating in a cell which does not ordinarily
hydroxylated prolines. The method comprises optimizing a nucleic acid
sequence for expression in the cell by substitution of codons preferred
by that cell for naturally occurring codons not preferred by the cell;
incorporating the nucleic acid sequence into the cell; and contacting the
cell with a hypertonic growth medium containing at least one amino acid,
selected from the group consisting of trans-4-hydroxyproline and 3-
hydroxyproline to allow at least one of the amino acids to be assimilated
into the cell and incorporated into the extracellular matrix protein. The


```
Db 561 GAGLQGVGPGKAGLPGPKDRDAGPKGADGSGKDGKGVRLTGPFGPAGAGDCK 620
Qy 421 GSGSGGAPGPTGARGADGDRGEPGPPGAPGAGPAGDGFPGAKGFGDAGKADAGPP 480
Db 621 GSGSGGAPGPTGARGADGDRGEPGPPGAPGAGPAGDGFPGAKGFGDAGKADAGPP 680
Qy 481 GAGAGAGGPGTGDYCAPGAKGARGAGGPPGATGPPGAGRVGPPGSDAGPPGPPGA 540
Db 681 GAGAGAGGPPGPIGNVAPGAKGARGAGGPPGATGPPGAGRVGPPGSDAGPPGPPGA 740
Qy 541 GKEG 544
Db 741 GKEG 744
```

```
RESULT 3
ID AAY84403 standard; protein; 1058 AA.
AC AAY84403;
DT 12-JUL-2000 (first entry)
XX Amino acid sequence of human type 1 (alpha) collagen polypeptide.
```

```
XX Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
KW collagen; mussel adhesive protein; bioadhesive.
OS Homo sapiens.
XX WO200014201-A1.
XX 16-MAR-2000.
XX 07-SEP-1999; 99WO-US020462.
XX 09-SEP-1998; 98US-0099652P.
XX (USSU) US SURGICAL CORP.
XX (PAOL) PAOLELLA D N.
XX (GRUS) GRUSKIN E A.
XX (BUEC) BUECHTER D D.
XX Paolella DN, Gruskin EA, Buechter DD;
XX WPI; 2000-271051/23.
XX N-PSDB; AA299843.
XX Incorporating non-natural amino acid into polypeptide, useful e.g. for
XX production of bioadhesives, by epoxidation or substitution of
XX dehydroproline residues.
XX Disclosure; Fig 6; 66pp; English.
```

```
XX The present sequence represents a human type 1 (alpha) collagen protein.
XX Peptides derived from the protein were used to demonstrate incorporation
XX of 3,4-dehydro-L-proline into the peptide, using the method of the
XX invention. The specification describes a method for the incorporation of
XX non-natural amino acid into a polypeptide. The method comprises reacting
XX at least one 3,4-dehydroproline residue in the polypeptide with an
XX epoxidation reagent from a polypeptide containing at least one 3,4-
XX epoxyproline residue. The method is used for studying the effects of non-
XX natural amino acids on structure and function of polypeptides. The method
XX is also useful for commercial production of collagen or mussel adhesive
XX proteins (which are useful as bioadhesives), and for incorporating a wide
XX variety of groups, including therapeutic ligands and biological probes,
XX into polypeptides
XX Sequence:1058 AA;
```

```
Query Match 96.6%; Score 2966; DB 3; Length 1058;
Best Local Similarity 95.6%; Pred. No. 2.7e-169;
```

```
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GSGSGGVRGEPGPPGAPGAGPAGDGFPGAKGFGDAGKADAGPPGARGSGPGE 60
Db 202 GSGSGGVRGEPGPPGAPGAGPAGDGFPGAKGFGDAGKADAGPPGARGSGPGE 261
Qy 61 GPGGPPGPKDSDGEPGAPGSKGDTGAKGBPGPVGVGPPGAPGEEGKPGARGPPTGLP 120
Db 262 GPGGPPGPKDSDGEPGAPGSKGDTGAKGBPGPVGVGPPGAPGEEGKPGARGPPTGLP 321
Qy 121 GPPGERGGPSRFPFGADGVAGPKGAPAGRGSGPAGPKGSGPGEAGRPGEAGLPGAKGLT 180
Db 322 GPPGERGGPSRFPFGADGVAGPKGAPAGRGSGPAGPKGSGPGEAGRPGEAGLPGAKGLT 381
Qy 181 GSPGSGPFGDKTGPFGPAGDGRFPFGPPGARGGAGVGMGFPKGAAGBPGKAGRGVP 240
Db 382 GSPGSGPFGDKTGPFGPAGDGRFPFGPPGARGGAGVGMGFPKGAAGBPGKAGRGVP 441
Qy 241 GPPGAVGPAGKDGAEAGGPPGAPGAGRGSGPAGSGPGEAGLPGAGKPGGEE 300
Db 442 GPPGAVGPAGKDGAEAGGPPGAPGAGRGSGPAGSGPGEAGLPGAGKPGGEE 501
Qy 301 GVPQDILGAPGSGARGEPGPPGARGGVEGPPGAPGADGAPGDDGAKGADAGAPGSE 360
Db 502 GVPQDILGAPGSGARGEPGPPGARGGVEGPPGAPGADGAPGDDGAKGADAGAPGSE 561
Qy 361 GARGLEGMPGERGAAGLPGPKGDRGADGPKGADGSGFKDGVRLTGPFGPPGAPAGDX 420
Db 562 GARGLEGMPGERGAAGLPGPKGDRGADGPKGADGSGFKDGVRLTGPFGPPGAPAGDX 621
Qy 421 GSGSGGAPGPTGARGADGDRGEPGPPGAPGAGPAGDGFPGAKGFGDAGKADAGPP 480
Db 622 GSGSGGAPGPTGARGADGDRGEPGPPGAPGAGPAGDGFPGAKGFGDAGKADAGPP 681
Qy 481 GPAGAGPAGPPIGDVQAPGAKGARGAGPAGGATGFPGAGRVGPPGSDAGPPGPPGA 540
Db 682 GPAGAGPAGPPIGNVQAPGAKGARGAGPAGGATGFPGAGRVGPPGSDAGPPGPPGA 741
Qy 541 GKEG 544
Db 742 GKEG 745
```

RESULT 4

AAR89472
ID AAR89472 standard; protein; 1107 AA.

XX AAR89472;
XX 01-OCT-1996 (first entry)
XX Collagen/decorin(aa46-93) fusion protein.

XX Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
XX bone formation; tissue repair; fusion protein.
XX Synthetic.

XX	Key	Location/Qualifiers
XX	Domain	1..1057
XX		/label= Collagen-IA
XX		/note= "collagen IA alpha-helical domain"
XX	Misc-difference	887
XX		/note= "unidentified amino acid"
XX	Misc-difference	890
XX		/note= "unidentified amino acid"
XX	Peptide	1058..1059
XX		/label= linker_peptide
XX	Domain	1060..1107
XX		/label= Decorin
XX		/note= "amino acids P46 to G93 of mature decorin"

CA2151547-A.

Best Local Similarity 95.6%; Pred. No. 2.8e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAGAGADPCADGEPKAGKADGAPGAGPPGARGSPGE 60
Db 201 GSEGPQGVGRGPPGPPGAGAGADPCADGEPKAGKADGAPGAGPPGARGSPGE 260
QY 61 GPGPPGPKGDSGFBGAFSGKGTCAKGEPPGVGVEGPPGAGEGKPGARGSPGPTGLP 120
Db 261 GPGPPGPKGDSGFBGAFSGKGTCAKGEPPGVGVEGPPGAGEGKPGARGSPGPTGLP 320
QY 121 GPPERGSGRGPAGADCVAGPKPAGERSGPPAGKSPGAGPPGAGLPGAKGLT 180
Db 321 GPPERGSGRGPAGADCVAGPKPAGERSGPPAGKSPGAGPPGAGLPGAKGLT 380
QY 181 GSPSGPDGKTGTGPPGAGEDRPPGPPGARGGAGVMPGPKGAAGPPGKAGRGVP 240
Db 381 GSPSGPDGKTGTGPPGAGEDRPPGPPGARGGAGVMPGPKGAAGPPGKAGRGVP 440
QY 241 GPPGAVGAGKDGEMAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 300
Db 441 GPPGAVGAGKDGEMAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 500
QY 301 GVPDGLGAPGSGARGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 360
Db 501 GVPDGLGAPGSGARGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 560
QY 361 GAPLEGMPBRGAAGLPKPKGDRDAGPKGADGSPGKGVGRGLTGTPTGPPGAGAGDK 420
Db 561 GAPLQGMFBRGAAGLPKPKGDRDAGPKGADGSPGKGVGRGLTGTPTGPPGAGAGDK 620
QY 421 GESGSGGAGPTGARGAGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 480
Db 621 GESGSGGAGPTGARGAGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 680
QY 481 GAPGAGPPGPTGVDGAPGAGKAGSAGPPGATGPPGAGGPPGAGGPPGAGGPPG 540
Db 681 GAPGAGPPGPTGVDGAPGAGKAGSAGPPGATGPPGAGGPPGAGGPPGAGGPPG 740
QY 541 GKEG 544
Db 741 GKEG 744
RESULT 6
ADE87050
ID ADE87050 standard; protein; 1161 AA.
XX AC ADE87050;
XX DT 29-JAN-2004 (first entry)
XX DE Human pancreatic cell protein sequence SeqID510.
XX KW neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
XX KW cancer death; cytostatic; vaccine; gene therapy;
XX KW non-cancerous pancreas disease; human.
XX OS Homo sapiens.
XX PN WO2003060145-A2.
XX PD 24-JUL-2003.
XX PF 19-DEC-2002; 2002WO-US040655.
XX PR 21-DEC-2001; 2001US-0342768P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Sun Y, Liu C;
XX DR WPI; 2003-587286/55.

DR N-PSDB; ADE87387.
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX Claim 12; SEQ ID NO 510; 635pp; English.
XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.
XX SQ Sequence 1161 AA;
Query Match 96.6%; Score 2966; DB 7; Length 1161;
Best Local Similarity 95.6%; Pred. No. 2.9e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGPVGRGPPGPPGAGAGADPCADGEPKAGKADGAPGAGPPGARGSPGE 60
Db 59 GSEGPVGRGPPGPPGAGAGADPCADGEPKAGKADGAPGAGPPGARGSPGE 118
QY 61 GPGPPGPKGDSGFBGAFSGKGTCAKGEPPGVGVEGPPGAGEGKPGARGSPGPTGLP 120
Db 119 GPGPPGPKGDSGFBGAFSGKGTCAKGEPPGVGVEGPPGAGEGKPGARGSPGPTGLP 178
QY 121 GPPERGSGRGPAGADCVAGPKPAGERSGPPAGKSPGAGPPGAGLPGAKGLT 180
Db 179 GPPERGSGRGPAGADCVAGPKPAGERSGPPAGKSPGAGPPGAGLPGAKGLT 238
QY 181 GSPSGPDGKTGTGPPGAGEDRPPGPPGARGGAGVMPGPKGAAGPPGKAGRGVP 240
Db 239 GSPSGPDGKTGTGPPGAGEDRPPGPPGARGGAGVMPGPKGAAGPPGKAGRGVP 298
QY 241 GPPGAVGAGKDGEMAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 300
Db 299 GPPGAVGAGKDGEMAGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 358
QY 301 GVPDGLGAPGSGARGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 360
Db 359 GVPDGLGAPGSGARGGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 418
QY 361 GAPLEGMPBRGAAGLPKPKGDRDAGPKGADGSPGKGVGRGLTGTPTGPPGAGAGDK 420
Db 419 GAPLQGMFBRGAAGLPKPKGDRDAGPKGADGSPGKGVGRGLTGTPTGPPGAGAGDK 478
QY 421 GESGSGGAGPTGARGAGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 480
Db 479 GESGSGGAGPTGARGAGPPGPPGAGGPPGAGGPPGAGGPPGAGGPPGAGGPP 538
QY 481 GAPGAGPPGPTGVDGAPGAGKAGSAGPPGATGPPGAGGPPGAGGPPGAGGPPG 540
Db 539 GAPGAGPPGPTGVDGAPGAGKAGSAGPPGATGPPGAGGPPGAGGPPGAGGPPG 598
QY 541 GKEG 544
Db 599 GKEG 602
RESULT 7
AAR89469
ID AAR89469 standard; protein; 1169 AA.
XX AC AAR89469;
XX DT 01-OCT-1996 (first entry)

[illegible]

XX The specification describes a method for producing an extracellular
CC matrix protein or its fragment. The extracellular matrix protein is
CC capable of self aggregating in a cell which does not ordinarily
CC hydroxylated prolines. The method comprises optimising a nucleic acid
CC sequence for expression in the cell by substitution of codons preferred
CC by that cell for naturally occurring codons not preferred by the cell;
CC incorporating the nucleic acid sequence into the cell; and contacting the
CC cell with a hypertonic growth medium containing at least one amino acid,
CC selected from the group consisting of trans-4-hydroxyproline and 3-
CC hydroxyproline to allow at least one of the amino acids to be assimilated
CC into the cell and incorporated into the extracellular matrix protein. The
CC method may be used to make host cells assimilate and incorporate trans-4-
CC hydroxyproline into proteins. This is especially useful in the
CC recombinant production of proteins such as collagen, fibrinogen and
CC fibronectin whose ability to self aggregate and produce functional
CC proteins depends on the post translational hydroxylation of prolines. The
CC method is also useful in studying the structure and function of
CC polypeptides which do not normally contain trans-4-hydroxyproline. The
CC present sequence represents a chimeric collagen 1 (alpha1)/bone
CC morphogenic protein-2a (bmp-2b) protein, which may be produced using the
CC method of the invention
XX
SQ Sequence 1169 AA;
Query Match 96.6%; Score 2966; DB 3; Length 1169;
Best Local Similarity 95.6%; Pred. No. 2.9e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGDFGAGKADGAPGIAGAPFGPGARGSPGE 60
Db 201 GSEGPQGVGRGPPGPPGAGAGPAGNFGADGQCGAKGANGAPGIAGAPFGPGARGSPGQ 260
QY 61 GGGGPPPKGDSGEPAPKSGDGTGAKGEPGVGVGVEGPPGAGSEKPKGARGEPGTGLP 120
Db 261 GPGGPPPKGNGSEGPAPKSGDGTGAKGEPGVGVGVEGPPGAGSEKPKGARGEPGTGLP 320
QY 121 GPPGERGPGSGRFFPGADGVAGPKGPPAGRGSPGPPAGPKGSPGAEGRPGAEGLPGAKGLT 180
Db 321 GPPGERGPGSGRFFPGADGVAGPKGPPAGRGSPGPPAGPKGSPGAEGRPGAEGLPGAKGLT 380
QY 181 GSPGSPGPPDKTPTPPGPPAGEDGPPGPPGARGENGVMGPPGKAGGEPKAGRGVVP 240
Db 381 GSPGSPGPPDKTPTPPGPPAQDQGRPPGPPGARGQGVGPPGKAGGEPKAGRGVVP 440
QY 241 GPPGAVGPPKDGAEAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
Db 441 GPPGAVGPPKDGAEAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 500
QY 301 GVPEDLGAQPSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
Db 501 GVPEDLGAQPSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 560
QY 361 GAPGLEMPGRCGAAGLPKPKGDRGADGPKGADGSPKDGVRGLTGPFGPPGPPGAPGDK 420
Db 561 GAFGLQMPGRCGAAGLPKPKGDRGADGPKGADGSPKDGVRGLTGPFGPPGPPGAPGDK 620
QY 421 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db 621 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 680
QY 481 GPAGGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
Db 681 GPAGGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 740
QY 541 GKEG 544
Db 741 GKEG 744
RESULT 9
AAR89470
ID AAR89470 standard; protein; 1171 AA.

XX AAR89470;
AC 01-OCT-1996 (first entry)
DT Collagen/TGF-beta-1 fusion protein.
DE Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KW bone formation; tissue repair; fusion protein.
XX Synthetic.
OS
FH Key Location/Qualifiers
FT Domain 1..1057
FT /label= Collagen-IA
FT /note= "collagen IA alpha-helical domain"
FT Misc-difference 887
FT /note= "unidentified amino acid"
FT Misc-difference 890
FT /note= "unidentified amino acid"
FT Peptide 1058..1059
FT /label= linker_peptide
FT Domain 1060..1171
FT /label= TGF-beta-1
FT /note= "human mature TGF-beta-1"
XX CA2151547-A.
PN 11-DEC-1995.
XX 12-JUN-1995; 95CA-02151547.
PF 10-JUN-1994; 94US-00259263.
PR (US)) US SURGICAL CORP.
XX Gruskin EA, Espino P;
XX WPI; 1996-140144/15.
XX N-PSDB; AAT16516.
XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
PT - and cellular regulatory factor domain, partic. useful as osteogenic
PT agents, also related vectors, transformed cells and chimaeric proteins.
XX Disclosure; Fig 6; 59pp; English.
XX A fusion protein (AAR89470) comprises the alpha-helical region of human
CC collagen I(a) linked to the human mature transforming growth factor beta-
CC 1 (TGF-beta-1). It can be expressed in Escherichia coli transformants
CC carrying a vector incorporating a chimeric gene (AAT16516) coding for the
CC fusion. The TGF-beta-moiety increases efficacy of the body's normal soft
CC tissue repair response and also induces osteogenesis. The collagen moiety
CC provides an integral substratum or scaffolding for the TGF and cells
CC involved in reconstruction and growth. The fusion protein provides
CC sustained release and delivery of TGF-beta-1 to a target tissue
XX
SQ Sequence 1171 AA;
Query Match 96.6%; Score 2966; DB 2; Length 1171;
Best Local Similarity 95.6%; Pred. No. 2.9e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGDFGAGKADGAPGIAGAPFGPGARGSPGE 60
Db 201 GSEGPQGVGRGPPGPPGAGAGPAGNFGADGQCGAKGANGAPGIAGAPFGPGARGSPGQ 260
QY 61 GPGGPPPKGDSGEPGAPKSGDGTGAKGEPGVGVGVEGPPGAGSEKPKGARGEPGTGLP 120
Db 261 GPGGPPPKGNGSEGPAPKSGDGTGAKGEPGVGVGVEGPPGAGSEKPKGARGEPGTGLP 320
QY 121 GPPGERGPGSGRFFPGADGVAGPKGPPAGRGSPGPPAGPKGSPGAEGRPGAEGLPGAKGLT 180
XX

[illegible]

RESULT 10

AA84538
ID AA84538 standard; protein; 1171 AA.

DT 25-JUL-2000 (first entry)

DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.

Extracellular matrix protein; self aggregation; hydroxylated proline; trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production; collagen; fibrinogen; fibronectin; post translational hydroxylation; ss. transforming growth factor-beta1; TGF-beta1; chimera.

xx	
OS	Homo sapiens.
OS	Unidentified.
OS	Chimeric.

AA				
FH	Key	Location/Qualifiers		
FT	Misc-difference	858	/note= "gly encoded by GCT"	
FT				

XX PN EP992586-A2.

XX
PD
12-APR-2000.

XX
PF 07-OCT-1999: 99EP-00119184.

XX
PR 09-OCT-1998: 98US-00169768.XX
PA (USSU) US SURGICAL CORP.

PI Gruskin EA, Buechter DD, Zhang G, Connolly K; XX

XX
DR WPI: 2000-259138/23.

DR N-PSDB; AAA12498.
XX

PT Production of extracellular matrix proteins containing 4-trans-hydroxyproline results in native self aggregating proteins, useful on medical implants.

RESULT 11
ADE87057

Claim 23; Fig 15; 260pp; English.

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents chimeric collagen 1 (alpha1)/transforming growth factor-beta1 (TGF-beta1) protein, which may be produced using the method of the invention.

Sequence 1171 AA;

Query Match 96.6%; Score 2966; DB 3; Length 1171;

Best Local Similarity 95.6%; Pred. No. 2.9e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

[illegible]

RESULT 11
ADE87057

QY	1	GSEGEVGRGEPPGPPGAGAGPADPGADGEPGAKGADGAPCIAGAPFPFGARGSPGE	60
Db	362	GSEGPCVGRGEPPGPPGAGAGPAGNPGADQCPGAKGANGAPGIAGAPFPFGARGSPGQ	421
QY	61	GPQGPCKGDSGEGPAGSGKDTCAKGEPPVGVGPPGPAEGEKGPAEGEPPTGLP	120
Db	422	GPQGPCKGDSGEGPAGSGKDTCAKGEPPVGVGPPGPAEGEKGPAEGEPPTGLP	481
QY	121	GPFGGPGSGRFFGADGVAQKPGPAGRGSPGAPGKSGPGEAGRPGEAGLPGAAGLT	180
Db	482	GPFGGPGSGRFFGADGVAQKPGPAGRGSPGAPGKSGPGEAGRPGEAGLPGAAGLT	541
QY	181	GSFGSPGDPCKTPPGPAGEDGRPGPPPGPARGAEGVMGPPGPKGAAGEPGKAGRGVP	240
Db	542	GSFGSPGDPCKTPPGPAGEDGRPGPPPGPARGAEGVMGPPGPKGAAGEPGKAGRGVP	601
QY	241	GPFGVAPGAGKGEAGAGPPGPPGAPGAGRGEEGPPGPFEGILPGAPGPPGAGKFGEE	300
Db	602	GPFGVAPGAGKGEAGAGPPGPPGAPGAGRGEEGPPGPFEGILPGAPGPPGAGKFGEE	661
QY	301	GVPGDLGAPGSGARGEPFGPGERVVEGPPGPPGADGAPDDGAKGDAGAPGAPGSE	360
Db	662	GVPGDLGAPGSGARGEPFGPGERVVEGPPGPPGADGAPDDGAKGDAGAPGAPGSE	721
QY	361	GAPGLEMPGERGAAGLPGKCDRGDAGPKGADGSPGKGVGRGILGPIGPPGAPAGPKD	420
Db	722	GAPGLEMPGERGAAGLPGKCDRGDAGPKGADGSPGKGVGRGILGPIGPPGAPAGPKD	781
QY	421	GESGSPGAPGTCARGAPDGRGEPGPPGAPGPPGADGEPGAKGEPFGDAGKGDAGPP	480
Db	782	GESGSPGAPGTCARGAPDGRGEPGPPGAPGPPGADGEPGAKGEPFGDAGKGDAGPP	841
QY	481	GPAGPAGPPTGDIYCAPKAGKAGSAPPGCATGFCGACRGVGPSPGSDAGPPGPPGA	540
Db	842	GPAGPAGPPTGDIYCAPKAGKAGSAPPGCATGFCGACRGVGPSPGSDAGPPGPPGA	901
QY	541	GKEG 544	
Db	902	GKEG 905	
RESULT 13			
AAR89471	ID	AAR89471 standard; protein; 1388 AA.	
XX	AC	AAR89471;	
XX	DT	01-OCT-1996 (first entry)	
XX	DE	Collagen/decorin fusion protein.	
XX	KW	Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;	
XX	XX	bone formation; tissue repair; fusion protein.	
XX	OS	Synthetic.	
XX	Key	Location/Qualifiers	
FT	Domain	1..1057	
FT		/label= Collagen-IA	
FT		/note= "collagen IA alpha-helical domain"	
FT	Misc-difference	887	
FT		/note= "unidentified amino acid"	
FT	Misc-difference	890	
FT		/note= "unidentified amino acid"	
FT	Peptide	1058..1059	
FT		/label= Linker_peptide	
FT	Domain	1060..1388	
FT		/label= Decorin	
XX	XX		
PN	CA2151547-A.		
XX	11-DEC-1995.		
PD			

cc cell (DACC) clones disclosed. More particularly, the method relates to
cc these polypeptides stimulating mesenchymal cell growth and/or division
cc and to transfecting these cells and chondrocytes with vectors carrying
cc the genes of these polypeptides capable of stimulating chondrogenesis,
cc osteogenesis, growth, repair, regeneration and/or restoration of the
cc extracellular matrix. The chondrocytes selectively express genes required
cc to form a cartilaginous matrix. The DACC polypeptides and polynucleotides
cc are useful for identifying an agent that modulates the activity of the
cc polypeptide, for stimulating mesenchymal cell growth and/or division by
cc exposing animal mesenchymal cells to conditioned media or its active
cc fraction, obtained from deer antler cartilage cells, for inhibiting cell
cc growth and/or division by inserting into an animal cell, a compound which
cc inhibits the translation of the polynucleotide encoding the DACC. The
cc method and the polypeptides are useful for stimulating mesenchymal cell
cc growth and/or division or for stimulating chondrogenesis, cartilage, disc
cc or connective tissue growth, repair, regeneration and/or restoration in
cc an animal. The polynucleotides, polypeptides, agonists and antagonists
cc may be used in treatment modalities, specifically in gene therapy. The
cc polypeptides can be used as bait proteins in a two- or three-hybrid assay
cc to identify other proteins, which bind to or interact with the
cc polypeptide and are involved in modulating cell growth and/or division.
cc The sequences presented in ABG93923-ABG93948 are the proteins encoded by
cc the DACC cDNA clones
xx
SQ Sequence 1461 AA;

Query Match		96.6%	Score 2966;	DB 5;	Length 1461;
Best Local Similarity		95.6%	Pred. No. 3.5e-169;		
Matches 520;		Conservative 21;	Mismatches 3;	Indels 0;	Gaps 0;
Qy	1	GSEGEVGRGPPGPPGAGAGAGDGFAGDGEFGAKGADGAGIAGAPGPGARGSPGE	60		
Db	359	GSEGPQGVGRGPPGPPGAGAGAGDQPGAKGANGAGIAGAPGPGARGSPGPQ	418		
Qy	61	GPQGPGRKDSBEPGAPGSKGDTGAKGEPGVGVEGPPGAGEGKPGARGEPGTGLP	120		
Db	419	GPQGPFGKNGSBEPGAPGSKGDTGAKGEPGVGVEGPPGAGEGKRGARGEPGTGLP	478		
Qy	121	GPPGERGSGSRGPPFGADGVAGPKGAPAGERGSGPGAPGKSGPGEAGRPGEAGLP	180		
Db	479	GPPGERGSGSRGPPFGADGVAGPKGAPAGERGSGPGAPGKSGPGEAGRPGEAGLP	538		
Qy	181	GSPGSPGPDGKTGTPPGAGEDGRPPGPPGARGAGVWGFPCKNAGEPGKAGERGVP	240		
Db	539	GSPGSPGPDGKTGTPPGAGDGRPPGPPGARGAGVWGFPCKNAGEPGKAGERGVP	598		
Qy	241	GPPCAVGFAGKDGAGAGGPPGPPGAPAGERGEGEPAGSPGFEGLPGPAGPPGAGKPGEE	300		
Db	599	GPPCAVGFAGKDGAGAGGPPGPPGAPAGERGEGEPAGSPGFEGLPGPAGPPGAGKPGEE	658		
Qy	301	GVPDGLGAPGSPGARGSPGPPGVEGPPGPPGADGAPGDDGAKGADAGAPGAPGSE	360		
Db	659	GVPDGLGAPGSPGARGSPGPPGVEGPPGPPGADGAPGDDGAKGADAGAPGAPGSE	718		
Qy	361	GAPLEGMPGERRAAGLPKPKDRDAGPKGADGSPKGVRLTGTPIGPPGAPAGPDK	420		
Db	719	GAPLEGMPGERRAAGLPKPKDRDAGPKGADGSPKGVRLTGTPIGPPGAPAGPDK	778		
Qy	421	GESGSPGAPGTGARGAPGRRGEPGPPGAPGAPGADGEPGAKGEPGADGAGP	480		
Db	779	GESGSPGAPGTGARGAPGRRGEPGPPGAPGAPGADGEPGAKGEPGADGAGP	838		
Qy	481	GPAGPAGPPGPIGVGAPGARGAGSPGATGTPGCAAGRVGPPGSPGAGPPGPPGA	540		
Db	839	GPAGPAGPPGPIGVGAPGARGAGSPGATGTPGCAAGRVGPPGSPGAGPPGPPGA	898		
Qy	541	GKEG 544			
Db	899	GKEG 902			

Search completed: May 7, 2004, 16:28:20
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 16:27:10 ; Search time 23 Seconds
(without alignments)
1221.066 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGEVGRGEPGPPGAGA.....PQSGDAGPPGPPGAGKEG 544
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/iaa/PCUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/iaa/backfile1.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2966	96.6	1461	US-09-585-887-9	Sequence 9, Appli
2	2966	96.6	1461	US-09-585-887-9	Sequence 9, Appli
3	2966	96.6	1461	US-09-331-347C-21	Sequence 21, Appl
4	2857	93.1	822	US-09-219-849-49	Sequence 49, Appl
5	2842	92.6	1057	US-08-931-820-1	Sequence 1, Appli
6	2791.5	90.9	1341	US-08-963-825-18	Sequence 18, Appl
7	2791.5	90.9	1341	US-09-500-811-18	Sequence 18, Appl
8	2791.5	90.9	1341	US-09-570-573-18	Sequence 18, Appl
9	2791.5	90.9	1341	US-09-548-608-18	Sequence 18, Appl
10	2226	72.5	1017	US-08-468-996-10	Sequence 10, Appl
11	2202	71.7	1060	US-08-931-820-3	Sequence 3, Appli
12	2202	71.7	1418	US-08-963-825-20	Sequence 20, Appl
13	2202	71.7	1418	US-09-010-999-1	Sequence 1, Appli
14	2202	71.7	1418	US-09-500-811-20	Sequence 20, Appl
15	2202	71.7	1418	US-09-570-573-20	Sequence 20, Appl
16	2202	71.7	1418	US-09-548-608-20	Sequence 20, Appl
17	2192	71.4	1442	US-08-316-850-12	Sequence 12, Appl
18	2192	71.4	1442	PCT-US95-02251-12	Sequence 12, Appl
19	2156	70.2	595	US-09-219-849-48	Sequence 48, Appl
20	2156	70.2	595	US-09-219-849-50	Sequence 50, Appl
21	1901	61.9	1057	US-08-931-820-4	Sequence 4, Appli
22	1886	61.4	1024	US-08-931-820-2	Sequence 2, Appli
23	1886	61.4	1366	US-08-963-825-19	Sequence 19, Appl
24	1886	61.4	1366	US-08-963-825-19	Sequence 19, Appl
25	1886	61.4	1366	US-09-500-811-19	Sequence 19, Appl
26	1886	61.4	1366	US-09-570-573-19	Sequence 19, Appl
27	1885	61.4	1366	US-09-548-608-19	Sequence 19, Appl
28	1885	61.4	1366	US-09-585-887-10	Sequence 10, Appl

28	1885	61.4	1366	4	US-09-289-578-10	Sequence 10, Appl
29	1871	60.9	1078	3	US-08-963-825-21	Sequence 21, Appl
30	1871	60.9	1078	4	US-09-500-811-21	Sequence 21, Appl
31	1871	60.9	1078	4	US-09-570-573-21	Sequence 21, Appl
32	1871	60.9	1078	4	US-09-548-608-21	Sequence 21, Appl
33	1640.5	53.4	1064	1	US-08-642-255-62	Sequence 62, Appl
34	1600	52.1	720	3	US-09-219-849-4	Sequence 4, Appli
35	1600	52.1	777	1	US-08-642-255-53	Sequence 53, Appl
36	1522.5	49.6	492	4	US-08-468-996-12	Sequence 12, Appl
37	1502.5	48.9	1065	1	US-08-642-255-72	Sequence 72, Appl
38	1496.5	48.7	561	1	US-08-642-255-52	Sequence 52, Appl
39	1490.5	48.6	504	3	US-09-219-849-3	Sequence 3, Appli
40	1447	47.1	633	3	US-08-642-255-73	Sequence 73, Appl
41	1414.5	46.1	960	3	US-09-219-849-5	Sequence 5, Appli
42	1406.5	45.8	829	1	US-08-642-255-132	Sequence 132, App
43	1406.5	45.8	829	1	US-08-357-633A-53	Sequence 53, Appl
44	1406.5	45.8	837	1	US-08-175-155-68	Sequence 68, Appl
45	1406.5	45.8	837	1	US-08-477-509B-103	Sequence 103, App

ALIGNMENTS

RESULT 1

US-09-585-887-9

; Sequence 9, Application US/09585887

; Patent No. 6413742

; GENERAL INFORMATION:

; APPLICANT: Olsen, David R

; APPLICANT: Chang, Robert

; APPLICANT: McMullin, Hugh

; APPLICANT: Hitzeman, Ronald A.

; APPLICANT: Chisholm, George

; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND

; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT

; TITLE OF INVENTION: CELLS

; FILE REFERENCE: 225002030400

; CURRENT APPLICATION NUMBER: US/09/585,887

; CURRENT FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: 09/289,578

; PRIOR FILING DATE: 1999-04-09

; PRIOR APPLICATION NUMBER: 60/084,828

; PRIOR FILING DATE: 1998-05-08

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1461

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-585-887-9

Query Match 96.6%; Score 2966; DB 4; Length 1461;

Best Local Similarity 95.6%; Pred. No. 8.9e-177;

Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY	1	GSEGEVGRGEPGPPGAGAAGDPCADGEPGAKGADGAPGAGPFGGARGPSPGPE	60
DB	359	GSEGEVGRGEPGPPGAGAAGDPCADGEPGAKGADGAPGAGPFGGARGPSPGPE	418
QY	61	GPCGPPGPKGDSGEPCAPGSKGDTGAKGEPGVGVGPPGAGGEGKFGARGPPTGLP	120
DB	419	GPCGPPGPKGDSGEPCAPGSKGDTGAKGEPGVGVGPPGAGGEGKFGARGPPTGLP	478
QY	121	GPPEGGGGERGFFGADGVAGPKGPPGAGRGSPGAGRGPEAGLPGAKGLT	180
DB	479	GPPEGGGGERGFFGADGVAGPKGPPGAGRGSPGAGRGPEAGLPGAKGLT	538
QY	181	GSFGSPGPDGKTGPPGAGEGDRPGPPGARGGAGVMGPPGPKGAAGEGPKAGERGVP	240
DB	539	GSFGSPGPDGKTGPPGAGEGDRPGPPGARGGAGVMGPPGPKGAAGEGPKAGERGVP	598
QY	241	GPFGVAGPKGDKGAGGAGGPPGPPGAGEGDRPGPPGARGGAGVMGPPGPKGAAGEGPKAGERGVP	300

Query Match	96.6%;	Score	2366;	DB	4;	Length	1451;
Best Local Similarity	95.6%;	Pred.	No. 8.9e-177;				
Matches	520;	Conservative	21;	Mismatches	3;	Indels	0;
Gaps	0						
Qy	1	GSEGGVGEVGEPPGPPGACGAGPAGDPPGADGEPGAKGADGAPGIAGAPGPPGAGSGPE	60				
Db	359	GSEGGVGEVGEPPGPPGACGAGPAGDPPGADGEPGAKGADGAPGIAGAPGPPGAGSGPE	418				
Qy	61	GGCGPPGPKDGEPCAGSKDCTCAKGEPPGVGEGPPGACGEGKPGAREGGTGLP	120				
Db	419	GGGGPPGPKNGSGEPAPGSKGDTCAKGEPPGVGQGGPPGAGEGKRGAREGGTGLP	478				
Qy	121	GPPEGGGPGSGRFFPGADGVAGPKPAGSGSPGAPGKSGPGBAGPGBAGLPGAKGLT	180				
Db	479	GPPEGGGPGSGRFFPGADGVAGPKPAGSGSPGAPGKSGPGBAGPGBAGLPGAKGLT	538				
Qy	181	GSPGSGPPDGKTPGPCAGEDGSGPPGPPPGARGEAGVWGPFGKGAAGSPGKAGERGVP	240				
Db	539	GSPGSGPPDGKTPGPCAQDGRPPGPPPGARQAGVWGPFGPKGAAGSPGKAGERGVP	598				
Qy	241	GPGAVGPKGKGAEAGPPPGAPGAGERGEPAGSPGPEGLPGPAGPPGAGKPGEE	300				
Db	599	GPGAVGPKGKGAEAGPPPGAPGAGERGEPAGSPGPEGLPGPAGPPGAGKPGEE	458				

Qy	1	GSEBEGVGRBEPGPPGPAAGADPGADGEPGAKGADGAPGDIAGAPGPPGARGSPGE	60
		:	
		:	
Db	362	GSEGPQGVGRBEPGPPGPAAGAPGNFPGADQPGAKGANGAPGIAGAPGPPGARGSPGQ	421
		:	
		:	
Qy	61	PGPGPPGPKGDSBEPGAPGSGDGTAKGEPGVGVGPPGPPAGBEGPKGARGPPGTGLP	120
		:	
Db	422	PGPGPPGPKGNSBEPGAPGSGDGTAKGEPGVGVQPPGPPAGBEGPKGARGPPGTGLP	481
		:	
Qy	121	PGPPGGRGPGSGRFFGADGVAGPKGPAGERSGPPAGPKGSPGABGPPGAGLTPGAKGLT	180
		:	
Db	482	PGPPGGRGPGSGRFFGADGVAGPKGPAGERSGPPAGPKGSPGABGPPGAGLTPGAKGLT	541
		:	
Qy	181	PGSPGSPDDKTTGPPPCPAGEDGRPPGPPPCPARGAEGVNGFPGPKGAAGBPKGARGVP	240
		:	
Db	542	PGSPGSPDDKTTGPPPGAPQDGRFPGPPPGAPQAGVNGFPGPKGAAGBPKGARGVP	601
		:	
Qy	241	PGPGAVGPAKDGEAGAGSPPGPAGPAGERCCEGPAGSPGFEGPLPGPAGPPGABGKPGEE	300
		:	
Db	602	PGPGAVGPAKDGEAGAGQPPGPAGPAGERCCEGPAGSPGFEGPLPGPAGPPGABGKPGEE	661
		:	
Qy	301	GVPGDLGAPGSPGARGEPGPFGERGVEGPPGAPGPGADGAPGDDGAKGDAGAPGAPGE	360
		:	
Db	662	GVPGDLGAPGSPGARGERGPFGERGVQPPGPPGAPGFRGANGAPGNDGAKGDAGAPGAPGSG	721
		:	
Qy	361	GAPGLEMGDGERGAAGLPDPKDDRGDAGPKGADSGPKDGVGRGLTGPIGPPGPAGAPGDK	420
		:	
Db	722	GAPLGQMPGERGAAGLPKPKDGRDAGPKGADSGPKDGVGRGLTGPIGPPGPAGAPGDK	781
		:	

[illegible]

Query Match	92.6%;	Score 2842;	DB 3;	Length 1057;
Best Local Similarity	91.9%;	Prod. No. 3.3e-163;		
Matches 500;	Conservative 24;	Mismatches 20;	Indels 0;	Gaps 0;
Qy	1	GSEPGVGRGEPGPPGAGAGPAGDPPGADGEPGAKGAGPAGGAGPFGGARGPSGPE	60	
Db	201	GSEPGQVRGEPGPPGAGAGPAGDPPGADGEPGAKGAGPAGGAGPFGGARGPSGPQ	260	
Qy	61	GPGGPPPKDSDSEPGAPGSKDGTAKGEPGPVGGVGGPPGAGREGKPGARGEPPGTCLP	120	
Db	261	PGGFTGARLVGEPGAPGKSGENKGEPPGAGPQGGPPGSGEGKRPNGEAGSGAPP	320	
Qy	121	GPGERGGPSRGFPAGDGVAGFGKGPAGBERGSPGAPGKSGPAGRPGEAGLFGAKGLT	180	
Db	321	GPGLRGGPSRGFPAGDGVAGFGKGPAGBERGSPGAPGKSGPAGRPGEAGLFGAKGLT	380	
Qy	181	GSPGSPGDKCTGPPGAGEDRGPPGPPGARGAGVWVGPPGKGAAGBPGKAGERGVP	240	
Db	381	GSPGSPGDKCTGPPGAGEDRGPPGPPGARGAGVWVGPPGKGAAGBPGKAGERGVP	440	
Qy	241	GPPGAVGPAGKDGAGAGFPAGPAGBERGEEGPAAGSPGFEGLPGPAGPPGGEAKGPGE	300	
Db	441	GPPGAVGPAGKDGAGAGFPAGPAGBERGEEGPAAGSPGFEGLPGPAGPPGGEAKGPGE	500	

APPLICANT: Bonde, Martin
FILING DATE: 08/187,319
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/570,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

Query Match 90.9%; Score 2791.5; DB 4; Length 1341;
Best Local Similarity 87.6%; Pred. No. 5.5e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

QY 1 GSEGPVGRGPPGPPGAGAGPAGDPGADGE---PGAKGADGAPGIAGAPGPPGAR--- 54
DB 211 GASGPMGPPGPPGKGBBZAGKPGPPGZGARGLPGTAGLPGMKHGRGFS 270
QY 55 -----GPSGPEG-----PG---GPPGKGDGSGEPGAPGSGKDTGAKGPPGV 93
DB 271 GLBGAAGBAGAPGKZPGSGZBGAFCZMGPZGKNSGEPGAPGSGKDTGAKGPPGV 330
QY 94 GVEGPPGAGEGKPGARGEPGTLGPPGGERGGSGRPPGADGAVGPKGPPAGERGSP 153
DB 331 GVQGPDPGAGEGKPGARGEPGTLGPPGGERGGSGRPPGADGAVGPKGPPAGERGSP 390
QY 154 GPAGPKGSPGAGRPGEAGLPGAKGLTGSPPSGPDGKTGPPGAGDGRPPGPPGAR 213
DB 391 GPAGPKGSPGAGRPGEAGLPGAKGLTGSPPSGPDGKTGPPGAGDGRPPGPPGAR 450
QY 214 GEAGVMGPPGKGAAGEGPKAGERGVPPGAVGPKDGEAGAGPPGPPGAPAGERGEE 273
DB 451 GQAGVMGPPGKGAAGEGPKAGERGVPPGAVGPKDGEAGAGPPGPPGAPAGERGEE 510
QY 274 GPAGSPGFEGLPGPAGPGEAGKPGEEGVPGDLAGAPGSGARGPPGPPGPPGPPGA 333
DB 511 GPAGSPGFEGLPGPAGPGEAGKPGEEGVPGDLAGAPGSGARGPPGPPGPPGPPGA 570

RESULT 8

US-09-570-573-18

; Sequence 18* Application US/09570573

; Patent No. 6342361

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

[illegible]

RESULT 10
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; TITLE OF INVENTION: GLUCAGON
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468

```

; PRIOR FILING DATE: 1990-10-10
;
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: prt
; ORGANISM: Homo sapiens
US-08-468-996-10

```

```
Query Match      72.5%; Score 2226; DB 4; Length 1017;
Best Local Similarity 72.6%; Pred. No. 5.2e-131;
Matches 395; Conservative 39; Mismatches 110; Indels 0
```

[illegible]

```

1  RESULT 11
2  US-08-931-820-3
3  Sequence 3, Application US/08931820
4  Patent No. 6010863
5  GENERAL INFORMATION:
6  APPLICANT:
7  TITLE OF INVENTION: Assay for c
8  NUMBER OF SEQUENCES: 4
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-D
13 SOFTWARE: PatentIn Release #1
14 CURRENT APPLICATION DATA:
15 'APPLICATION NUMBER: US/08/931
16 FILING DATE:
17 CLASSIFICATION: 435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: EP 962025

```

```

/ FILING DATE:
/
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1050 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Homo sapiens
/ TISSUE TYPE: Collagen type II
/ US-08-931-820-3

```

```
Query Match      71.7%; Score 2202; DB 3; Length 1060;
Best Local Similarity 71.3%; Pred. No. 1.7e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
```

QY	1	GSBGPBGVGRGPPPPGAPAGAPDGPADGEPGAKGADGAPGAPGPPGARGSPGPE	60
Db	203	GPEGAQGRGEPCTGSPGSPAGASNGPTDGPAGKSGACAPGIAGAPGPPGRGPPDPQ	262
QY	61	GPGGPPGPKDSEPCGAPSKGDTGAKGPPGVGVGPPGAPGEEGKPGARGPPGPTGLP	120
Db	263	GATGPGJPGKGTQKGIAGFKGQQGKGPFPAGPQCAPGAPGEBKRGARGPPGVGPI	332
QY	121	GPPEGRGGPSRRFPDAGDVAGPGKGPAGRGSPGAPGKSGPGEAGRPGEAGLPGAKGLT	180
Db	323	GPPEGERGAPGNRGFPQDDGLAGPKGAPGERGPPSLAGPKGANGDPGRPEPGLPGARGLT	382
QY	181	GPSGSPGPGKGTGPPGPAGDEGRPGPPGPPGARGENGVMGFPDGPKGAAGPFGKAGERVFP	240
Db	383	GRPGDAGPQKVQSPGAPGEDRPGPPGQCARGPQGVNMGFPFGKANGPPFGKAGEKSLP	442
QY	241	GPPGAVGPAGKQGEACGAECPGPGAPGARGERGPPAGSPGFEGLLPGAPGPPGAGKPGEE	300
Db	443	GAPGLKGLPKQKGETGAEGLPPGPAGPAGERGEQCAPGPPSGFQGLPGLPPGPPGGGKRPDQ	502
QY	301	GVPPDGLACGPPSGARGEPFPGERGVGPPGPPGAPPCADGAPDDGAKGDAGAPGAPGSE	360
Db	503	GVPGEAGAPGLVGRGERGFPGERGSPGACQLGQGRGLPCTGDPKPGASGAPPGPAQ	562
QY	361	GAPGLEMPGERGAAGLPBPKGDRDAGPKGADGSPKGQVRLGTGPIGPBPAGAPGDK	420
Db	563	GPPLQCMTPGERCAAGIAGPKGDRGVGEBKGPBGAPKDGGRGLTGPIGPPGPAGANGEK	622
QY	421	GESGPPGAPPTCARGAPCDRGEPPGPPGAPGAPGAPCADGCEPKAGEPCGDAGAKGDAGPP	480
Db	623	GEVGPFGPAGSAGAPGERGETGPPGTSGIAGPPGADGQPPKAGEQQGAGQKGDAGAP	682
QY	481	GPAPGAPGPGPIDGVAPGAKGAGSAGPPGATGTFPGAAGRVPPGPPSGDAGPPGPPGPA	540
Db	683	GPQPPSGAPGPQGTGVTGPKARGAQGPFGATGTFPGNAGRVGPPGNSGNPFGPPGPPS	742
QY	541	GKGB 544	
Db	743	GKDG 746	

```

RESULT 12
US-08-963-825-20
Sequence 20, Application US/08963825
Patent No. 6110689
GENERAL INFORMATION:
APPLICANT: Ovist, Martin
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for
TITLE OF INVENTION: in Body Fluid
TITLE OF INVENTION: Method and U
TITLE OF INVENTION: Disorders As
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESS: Darby & Darby PC

```

STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/963,825
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/187,319
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20
Query Match 71.7%; Score 2202; DB 3; Length 1418;
Best Local Similarity 71.9%; Pred. No. 2.1e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPGGPPGAGAGAGDPCADGEPGAKGADGAGTAGAGPPGARGSGPE 60
DB 315 GPEAGQGRGPGTGTGPGAGAGNPTDIPGAKGAGAGTAGAGPPGPPGPPDQ 374
QY 61 GPGGPPGKDGSGPFGAPSGDGTAKGEPGVGVGPPGPPGAGBEGKPGARGPFGTGLP 120
DB 375 GATGFLGPKGTGKPGTAGFKGEQPKGPGAGPQAGAPGAGEGKRGARGEPGVGPI 434
QY 121 GPPGGRGPGSGRPGADGVAGPKGAPAGERSGPGAPGKSPGAPGEGAGLPGAKGLT 180
DB 435 GPPGGRGAPNRGPPGQDGLAGPKGAPGERGPGSLAGPKGANGDPGPPGPPGARGLT 494
QY 181 GSPGSGPDGKTGPPGAGEDRPPGPPGPPGARGGAGVGMGPPGKGAAGFPKGAGRGVP 240
DB 495 GRPDAGQGVKVGSGAPGDEGRPPGPPGPPGARGQPGVMGPPGPKGANGEPGKAGEKGLP 554
QY 241 GPPGAVGPPAGKDGAGAGPPGPPGAPGAGERGEPGAPGPPGEGLPAGPPGAGKPGEE 300
DB 555 GAPGLRGLPGKGTGAGTGEPPGPPGAPGAGERGEPGAPGPPGPPGPPGPPGPPGPPGPPG 614
QY 301 GVPDGLGAPGSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 615 GVPGEAGAPGLVGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 674
QY 361 GAPGLEMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
DB 675 GPPGLQGNPGRGAGTAGKDRGVGKPEGAPGKDGGRGLTGIGIPPPGPPGAGANGEK 734
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB 735 CEVFPFGPAGAGARGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 794

QY 481 GPAGPAGPPGPIGDVGPAGPAGKAGSGAGPAGATGPPGAGAGRVGPPGPPGSDAGPPGPPA 540
DB 795 GPQGESGAPGPGTGTGTPGKARGAQPPGATGPPGAGRVGPPGPPGPPGPPGPPGPPGPPG 854
QY 541 GKEG 544
DB 855 GKDG 858
RESULT 13
US-09-010-999-1
Sequence 1, Application US/09010999
Patent No. 6132976
GENERAL INFORMATION:
APPLICANT: Poole, Anthony R.
APPLICANT: Hollander, Anthony P.
APPLICANT: Billingshurst, R. C.
TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,999
FILING DATE: 22-JAN-1998
CLASSIFICATION: 4335
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/448,501
FILING DATE: 17-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,123
FILING DATE: 04-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 032931/0212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Human Type II Collagen
US-09-010-999-1
Query Match 71.7%; Score 2202; DB 3; Length 1418;
Best Local Similarity 71.9%; Pred. No. 2.1e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
QY 1 GSEGEVGRGPGGPPGAGAGAGDPCADGEPGAKGADGAGTAGAGPPGARGSGPE 60
DB 315 GPEAGQGRGPGTGTGPGAGAGNPTDIPGAKGAGAGTAGAGPPGPPGPPDQ 374
QY 61 GPGGPPGKDGSGPFGAPSGDGTAKGEPGVGVGPPGPPGAGBEGKPGARGPFGTGLP 120
DB 375 GATGFLGPKGTGKPGTAGFKGEQPKGPGAGPQAGAPGAGEGKRGARGEPGVGPI 434

```

QY 121 GPPGERGGPSRGGPCGADGAGKPGAGRGSPGAGPKGSGPGEAGRPCEAGLPCAKGLT 180
Db 435 GPPGERGAPNRRGFGQDGLAFKAPGERGSPGLAGKANGDPRFGEPLPARGLT 494
QY 181 GSPGSPGPKTGTPGPPAGEDGRPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 240
Db 495 GRPGDAGPQGVGSPGAPGEDGRPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 554
QY 241 GPPGAVGAPAGKGEAGGPPGAPGAPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 300
Db 555 GAPLRLGPKDGETGAGPPGAPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 614
QY 301 GVPDGLGAPGSPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 360
Db 615 GVPGEAGAPGLVPRGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 674
QY 361 GAPLRLGPKDGETGAGPPGAPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 420
Db 675 GPPGQGVGPPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 734
QY 421 GSPGSPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 480
Db 735 GVPGEAGAPGLVPRGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 794
QY 481 GPAGPAGPPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 540
Db 795 GPQGVGPPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 854
QY 541 GKEG 544
Db 855 GKDG 858

```

RESULT 14

```

US-09-500-811-20
; Sequence 20, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-09-500-811-20

```

Query Match 71.7%; Score 2202; DB 4; Length 1418;

Best Local Similarity 71.9%; Pred. No. 2.1e-129;

Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

```

QY 1 GSEGEVGRGPPGPPGAGAGGAGDPCADGEGKAGKAGDAGGAGGPPGARGSPGE 60
Db 315 GPEAGQAPRGPPGPPGAGAGGAGDPCADGEGKAGKAGDAGGAGGPPGARGSPGE 374
QY 61 GPGGPPGPPGPPGAGAGGAGDPCADGEGKAGKAGDAGGAGGPPGARGSPGE 120
Db 375 GATGFLGPKQGTGPPGAGAGGAGDPCADGEGKAGKAGDAGGAGGPPGARGSPGE 434
QY 121 GPPGERGPPGPPGAGAGGAGDPCADGEGKAGKAGDAGGAGGPPGARGSPGE 180
Db 435 GPPGERGPPGPPGAGAGGAGDPCADGEGKAGKAGDAGGAGGPPGARGSPGE 494
QY 181 GSPGSPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 240
Db 495 GRPGDAGPQGVGSPGAPGEDGRPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 554
QY 241 GPPGAVGAPKGEAGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 300
Db 555 GAPLRLGPKDGETGAGPPGAPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 614
QY 301 GVPDGLGAPGSPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 360
Db 615 GVPGEAGAPGLVPRGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 674
QY 361 GAPLRLGPKDGETGAGPPGAPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 420
Db 675 GPPGQGVGPPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 734
QY 421 GSPGSPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 480
Db 735 GVPGEAGAPGLVPRGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 794
QY 481 GPAGPAGPPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 540
Db 795 GPQGVGPPGPPGARGGPPGPPGARGGPPGPPGARGAGVNGFFGPKGAAGEPKAGRGVVP 854
QY 541 GKEG 544
Db 855 GKDG 858

```

RESULT 15

```

US-09-570-573-20
; Sequence 20, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York

```

COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/570,573
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/187,319
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: GOSORIS, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN -ALPHA 1 (II)
 US-09-570-573-20

Query Match 71.7%; Score 2202; DB 4; Length 1418;
 Best Local Similarity 71.9%; Pred. No. 2.1e-129;
 Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

Qy	1	GSSEPEGVREPPGPPGAGAGAGDGEPCAKGADGAPCIAGAPFPAGRGSGPE	60
Db	315	GPEAQGRGEPGTPGSPGASGNGPTDGIPEAKGSAGAGIAGAPFPGRGPPDQ	374
Qy	61	GPSPGPKDSEPPGAPGSKDTGAKGEPGVGVEPPGAGEGKPGARGEPGTGLP	120
Db	375	GATGPLEGKQGTGKGTAGKGEQKGEFPGAPGAPGAGEGKRGARGEPGVGPI	434
Qy	121	GPGERGSGSPGPPGADGVAGPKPAGERSFPAGPKSGPGEAGRPGEAGLPAGKLT	180
Db	435	GPPGRCAGPNRGPPGQDGLAGPKGAPGERGSPGLAGPKGANGDPGRPGSPGLPGARLT	494
Qy	181	GPSPGPDGKTGPPGAGEDRGPPGPPGARGGAGVMPGPKGNAGEPKAGRGVYP	240
Db	495	GRPDAGFQKVGSPGAPGSDGRPPGPPGQARGQGVMGFPGPKGANGEPGKAGEKGLP	554
Qy	241	GPPGAVGPAGKDGAGAGPPGPPGAPGAGERGEGPAGSPGEGLPAGPPGAGKPGEE	300
Db	555	GAPGLRGLPKDGTGAEPPGPPGAPGAGERGEQAGPGSGFQGLPGPPPPGEGKPGDQ	614
Qy	301	GVPGDLGAPGSPGARGPPGPPGGERGVGPPGPPGADGAPDDGAKGADGAPGAPSE	360
Db	615	GVPEAGAPGLVGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	674
Qy	361	GAPGLEMPGRRGAAGLPKPKDRDAGPKGADSGPKDGVRLTGTGPPGPPGAPGDK	420
Db	675	GPPGLQMPGRRGAAGIAGPKGDRGDVGEKGEAPKDGGRGLTGTGPPGPPGANGEX	734
Qy	421	GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	480
Db	735	GEVGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	794
Qy	481	GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	540
Db	795	GPQFSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP	854

Qy 541 GKEG 544
 Db 855 GKDG 858
 Search completed: May 7, 2004, 16:30:57
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 16:28:25 ; Search time 49 Seconds
(without alignments)
3081.557 Million cells

Title: US-10-658-989A-4
 Perfect score: 3070
 Sequence: 1 GSEGPEGVGEPPGPPGAGA.....PGPSGDAGPPPGPAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Published Applications_A3.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	2366	96.6	1464	12	US-09-318-715-261 Sequence 261, App
2	2366	96.6	1464	14	US-10-060-036-159 Sequence 159, App
3	2366	96.6	1464	14	US-10-171-311-36 Sequence 36, Appl
4	2366	96.6	1464	14	US-10-216-705-21 Sequence 21, Appl
5	2366	96.6	1464	14	US-10-149-353-2 Sequence 2, Appl
6	2366	96.6	1464	14	US-10-177-293-65 Sequence 65, Appl
7	2366	96.6	1464	14	US-10-301-823-28 Sequence 28, Appl
8	2366	96.6	1464	15	US-10-391-265-243 Sequence 243, App
9	2336	95.6	1463	15	US-10-402-085-2 Sequence 2, Appl
10	2336	95.6	1463	15	US-10-402-072A-2 Sequence 2, Appl
11	2857	93.1	822	15	US-10-342-331-49 Sequence 49, Appl
12	2816.5 ^u	91.7	1449	15	US-10-402-085-8 Sequence 8, Appl
13	2816.5	91.7	1449	15	US-10-402-072A-8 Sequence 8, Appl
14	2791.5	90.9	1341	14	US-10-058-124-18 Sequence 18, Appl
15	2209	72.0	1014	14	US-10-194-441A-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1

```

US-09-918-715-261
; Sequence 261, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 261
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-261

```

Query Match	96.6%; Score 2956; DB 12; Length 1464;
Best Local Similarity	95.6%; Pred. No. 5.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;	
QY	1 GSEPGVGRGEPGPFGAAGAPGDAGDFGAKGADGAPGIAGAPFPFARGSGSPE 60
	: : : : : :
Dd	362 GSEPGCVGRGEPGPFGAAGPAGNFDGDPFGKANGAPGIAGAPFPFARGSGSQ 421
	: : : : : :
QY	61 GPGGPPRGKDGEPCAPAGSKDGTAKGPGPVGVGPPGPAAGEEKGARGEGPTGLP 120
	: : : : : :
Dd	422 GRGPRGPKNGSGPCAPAGSKDGTAKGPGPVGVGPPGPAAGEEKGARGEGPTGLP 481

QY 301 GVPDGLGAPGSGARGGPPGGRGVGPPGPPGAGPPGADGAPGDDGAKGDAGAPGSGE 360
Db 662 GVPDGLGAPGSGARGGPPGGRGVGPPGPPGAGPPGADGAPGDDGAKGDAGAPGSGE 721
QY 361 GAPLEGMPGEGRAAGLPGPKGRGDAGPKGADGSPKDGVRGLTGP1GPPGPPGAPGDK 420
Db 722 GARGLQMPGEGRAAGLPGPKGRGDAGPKGADGSPKDGVRGLTGP1GPPGPPGAPGDK 781
QY 421 GESGPPGAPGTPGARGAPGDRGPPGPPGAGPPGADGAPGDDGAKGDAGPP 480
Db 782 GESGPPGAPGTPGARGAPGDRGPPGPPGAGPPGADGAPGDDGAKGDAGPP 841
QY 481 GPAGPAGPPGPIGDNVGAAPGAKGARGSGNPPGATGFFCAAGRVGPPGSGNAGPPGPPGPA 540
Db 842 GPAGPAGPPGPIGDNVGAAPGAKGARGSGNPPGATGFFCAAGRVGPPGSGNAGPPGPPGPA 901
QY 541 GKEG 544
Db 902 GKEG 905

RESULT 6

US-10-177-293-65
; Sequence 65, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Giatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Marjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Martens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/239,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-65

Query Match 96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 5.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGPEGVGRGPPGPPGAGAGPAGDPGADGEPGAKGADGAPGTAGAPGPPGARGPSGPE 60
Db 362 GSEGPPQGVGRGPPGPPGAGAGPAGDPGADGEPGAKGADGAPGTAGAPGPPGARGPSGPE 421
QY 61 GPGGPPGPKGDSGPPGAGSGKGTGAKGEPGVPVEGPPGPPGAGGEPGKAGGEPGPTGLP 120
Db 422 GPGGPPGPKGNSGPPGAGSGKGTGAKGEPGVPVQGGPPGPPGAGGEPGKAGGEPGPTGLP 481
QY 121 GPPGERGPPGSRGPPGADGVAGPKGPAERKSGPPGAGPKGSPGABRPGGAGLFGAKGLT 180
Db 482 GPPGERGPPGSRGPPGADGVAGPKGPAERKSGPPGAGPKGSPGABRPGGAGLFGAKGLT 541
QY 181 GSPSPGPDGKGTGPPGAGEDRGPPGPPGARGGAGVMGPPGPKGAAGBPGKAGRGVP 240
Db 542 GSPSPGPDGKGTGPPGAGEDRGPPGPPGARGGAGVMGPPGPKGAAGBPGKAGRGVP 601
QY 241 GPPGAVGPAGKDGAGAGGPPGPPGAGPPGAGPPGAGPPGAGPPGAGPPGAGPPGAGPP 300
Db 602 GPPGAVGPAGKDGAGAGGPPGPPGAGPPGAGPPGAGPPGAGPPGAGPPGAGPPGAGPP 661
QY 301 GVPDGLGAPGSGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 662 GVPDGLGAPGSGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 721
QY 361 GAPLEGMPGEGRAAGLPGPKGRGDAGPKGADGSPKDGVRGLTGP1GPPGPPGAPGDK 420
Db 722 GARGLQMPGEGRAAGLPGPKGRGDAGPKGADGSPKDGVRGLTGP1GPPGPPGAPGDK 781
QY 421 GESGPPGAPGTPGARGAPGDRGPPGPPGAGPPGADGAPGDDGAKGDAGPP 480
Db 782 GESGPPGAPGTPGARGAPGDRGPPGPPGAGPPGADGAPGDDGAKGDAGPP 841
QY 481 GPAGPAGPPGPIGDNVGAAPGAKGARGSGNPPGATGFFCAAGRVGPPGSGNAGPPGPPGPA 540
Db 842 GPAGPAGPPGPIGDNVGAAPGAKGARGSGNPPGATGFFCAAGRVGPPGSGNAGPPGPPGPA 901
QY 541 GKEG 544
Db 902 GKEG 905

RESULT 7

US-10-301-822-28
; Sequence 28, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guilleminette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF COLON CANCER
; FILE REFERENCE: MM01-029F2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-28

```

RESULT 9
US-10-402-089-2
; Sequence 2, Application US/10402089
; Publication NO. US2004000563A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Bell, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Sealey, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 2
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-089-2

```

Query Match 95.6%; Score 2936; DB 15; Length 1463;
Best Local Similarity 94.1%; Pred. No. 1.9e-144;
Matches 512; Conservative 25; Mismatches 7; Indels 0; Gaps 0

```

RESULT 8
US-10-291-265-243
; Sequence 243, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-243

```



```

RESULT 12
US-10-402-089-8
; Sequence 8, Application US/10402089
; Publication NO. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 8
; LENGTH: 1449
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-8

```

121	QY	DB	180
484	DB	543	
181	QY	240	
544	DB	603	
241	QY	300	
604	DB	663	
301	QY	360	
664	DB	723	
361	QY	420	
724	DB	783	
421	QY	480	
784	DB	829	
481	QY	540	
830	DB	886	
541	QY		
887	DB		

```

RESULT 13
US-10-402-072A-8
; Sequence 8, Application US/10402072A
; Publication NO. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcus P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 1449
; TYPE: PRN
; ORGANISM: Sus scrofa
US-10-402-072A-8

```

	Query Match	91.7%	Score 2816.5	DB 15	Length 1449
	Best Local Similarity	91.0%	Pred. No. 2.9e-138		
	Matches 495	Conservative 22	Mismatches 10	Indels 17	Gaps 1
QY	1	GSEGEYVRGEPGPPGACAAAGPAGDGGADGGPGAKGADGAGIAGAPGPGARGGSGPGE	60		
DB	364	GSEGPQGVRGEPGPPGAGAGPAGNPGADGGPGKGANGAGIAGAPGPGARGGSGPQ	423		
QY	61	GFSGPPGPKGDSGEPCAPGSKDGTCAKGPPGVGVGPPGPGAGEEKGARGEPGTTGLP	120		
DB	424	GFSGPPGPKNGSEGFAPGSKDGTCAKGEPGTGTGQPPGAGEEKGARGEPGAGLP	483		
QY	121	GPPGENGPGSGRFFGADGVAGPKGAPAGRGSGPAGPKGSGPGEAGLPAGAKGLT	180		
DB	484	GPPGRRGGPSGRFFGADGVAGPKGAPAGRGSGPAGPKGSGPGEAGLPAGAKGLT	543		

QY 181 GSPGSGPQDKTGPAGAGDGRPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 240
Db 544 GSPGSGPQDKTGPAGAGDGRPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 603
QY 241 GPCAVGPAGKDCAGAGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 300
Db 604 GPCAVGPAGKDCAGAGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 663
QY 301 GVPGLDAGPSPGARGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 360
Db 664 GVPGLDAGPSPGARGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 723
QY 361 GARGLEMPGARGAGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 420
Db 724 GARGLEMPGARGAGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 783
QY 421 GESGPPGARGAGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 480
Db 784 GESGPPGARGAGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 829
QY 481 GPPAGPAGPPGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 540
Db 830 GPPAGPAGPPGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 886
QY 541 GKEG 544
Db 887 GKEG 890

RESULT 14
US-10-058-124-18
; Sequence 18, Application US/10058124
; Publication No. US20030119058A1
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; in Body Fluids, A Test Kit and Means for Carrying Out the
; Method and Use of the Method to Diagnose the Presence of
; Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/058,124
; FILING DATE: 29-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/570,573
; FILING DATE: 2002-MAY-12
; APPLICATION NUMBER: 08/187,319
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Ada C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

Query Match 90.9%; Score 2791.5; DB 14; Length 1341;
Best Local Similarity 87.6%; Pred. No. 5.3e-137;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

QY 1 GSEPGVGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 54
Db 211 GASGPMGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 270
QY 55 -----GSPGPEG-----PG--GPPGPKGDSGEPGAPGSGKDTGAKGEGPV 93
Db 271 GLBAGKCBAGPAGPKGZFGSPGZBGAFGZMGPPGPKGNSGEPGAPGSGKDTGAKGEGPV 330
QY 94 QVGGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 153
Db 331 QVGGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 390
QY 154 GPAGPKGSPGARGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 213
Db 391 GPAGPKGSPGARGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 450
QY 214 GEAGVMGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 273
Db 451 QGAGVMGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 510
QY 274 GPAGSPGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 333
Db 511 GPAGSPGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 570
QY 334 GPPGADGAPGDDGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 393
Db 571 GPRGANGAPGNDGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 630
QY 394 GSPGKGVRLTGTPIGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 453
Db 631 GSPGKGVRLTGTPIGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 690
QY 454 GPPGADGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 513
Db 691 GPPGADGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 750
QY 514 GPPGAGRVGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 544
Db 751 GPPGAGRVGPPGPPGARGAGRGPPGPPGARGAGRGPPGPPGARGAGVGMGFPKGAAGPGKAGRGVP 781

RESULT 15
US-10-194-441A-1
; Sequence 1, Application US/10194441A
; Publication No. US20030148944A1
; GENERAL INFORMATION:
; APPLICANT: Holmdahl, Rikard
; APPLICANT: Engstrom, Jan Ake
; APPLICANT: Kihlberg, Jan
; APPLICANT: Burkhardt, Harald
; TITLE OF INVENTION: TRIPLE POLYPEPTIDE COMPLEXES
; FILE REFERENCE: 11145-010001
; CURRENT APPLICATION NUMBER: US/10/194,441A
; CURRENT FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: US 60/305,048
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1

Search completed: May 7, 2004, 16:31:58
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 16:25:44 ; Search time 20 Seconds
(without alignments)
2616.410 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGPGEVGRGPPGPAGAA.....PGPSGDAGPPGPAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2966	96.6	1464	1 CGHU1S	collagen alpha 1(I)
2	2857	93.1	1453	2 S21626	collagen alpha 1(I)
3	2775	90.4	1042	1 CGGH1S	collagen alpha 1(I)
4	2220	72.3	1418	2 T49467	collagen alpha 1(I)
5	2202	71.7	1487	1 CGHU6C	collagen alpha 1(I)
6	2189	71.3	1419	2 A41182	collagen alpha 1(I)
7	2189	71.3	1487	2 B41182	collagen alpha 1(I)
8	2154	70.2	671	1 CGRT1S	collagen alpha 1(I)
9	2137	69.9	1486	2 B40333	collagen alpha 1(I)
10	2131	69.4	1492	2 A40333	collagen alpha 1(I)
11	2076.5	67.6	779	1 CGS01S	collagen alpha 1(I)
12	1921.5	62.6	1496	1 CGHU2V	collagen alpha 2(I)
13	1915	62.4	1373	1 A43291	collagen alpha 2(I)
14	1910	62.2	1497	2 I49607	procollagen type V
15	1908	62.1	1464	2 S59856	collagen alpha 1(I)
16	1905.5	62.1	1049	1 CGBO7S	collagen alpha 1(I)
17	1901	61.9	1466	1 CGHU7L	collagen alpha 1(I)
18	1886	61.4	1366	1 CGHU2S	collagen alpha 2(I)
19	1881	61.3	886	2 I50694	collagen alpha 1(I)
20	1550.5	50.5	1414	1 S23809	collagen alpha 2(I)
21	1457.5	47.5	1838	1 CGHU1E	collagen alpha 1(I)
22	1444	47.0	1806	1 CGHU1E	collagen alpha 1(I)
23	1432.5	46.7	1843	2 S18803	collagen alpha 1(I)
24	1418.5	46.2	1027	2 S28774	collagen alpha 1(I)
25	1413	46.0	964	1 CGGH2S	collagen alpha 2(I)
26	1411.5	46.0	1024	2 S18251	collagen alpha 1(I)
27	1400	45.6	888	2 S28791	collagen alpha 1(I)
28	1399.5	45.6	1546	1 CGHU2E	collagen alpha 2(I)
29	1376	44.8	1691	1 S22917	collagen alpha 5(I)

RESULT 1
CGHU1S
collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 31-Dec-2000
C/Accession: I60114; S01143; A93335; I55254; A39943; I55237; S09400; B90567; S15269; A29439; I53466; A02852; I37247
R/D/Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J
Gene 67, 105-115, 1988
A>Title: Complete nucleotide sequence of the region encompassing the first twenty-five A:Reference number: I60114; MUID:88329734; PMID:2843432
A:Accession: I60114
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369, 'L', 371-589 <DAL>
A:CROSS-references: GB:M20789; NID:G179593; PIDN:AAB59373.1; PID:G179594
R/Tomp, G.; Kuvantemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Proc Biochem. J. 253, 919-922, 1988
A>Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human A:Reference number: S01143; MUID:89025644; PMID:3178743
A:Accession: S01143
A:Molecule type: mRNA
A:Residues: 1-472 <TRO>
A:CROSS-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NI A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
R/Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C. Nature 310, 337-340, 1984
A>Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation o A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, 'Q', 60-181 <CHU>
A:CROSS-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658
R/Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W J. Biol. Chem. 262, 15151-15157, 1987
A>Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enl A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A>Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:CROSS-references: GB:J02829; NID:G180387; PIDN:AA51993.1; PID:G180388
R/Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Geinias, R.E. Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A>Title: Regulatory elements in the first intron contribute to transcriptional control A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:CROSS-references: GB:J03559; NID:G180876; PIDN:AA52052.1; PID:G553238
R/Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

ALIGNMENTS

collagen alpha 1(V)
collagen alpha 2(I)
type VII collagen
collagen alpha 4(I)
collagen alpha 1(X)
collagen alpha 1(I)
collagen alpha 3(I)
collagen alpha 1(V)
collagen alpha 1(I)
hypothetical prote
collagen alpha 1(I)
collagen alpha 2(I)
collagen COLF1 - f
collagen alpha 2(I)
collagen alpha 5(I)
collagen alpha 1(I)
collagen alpha 1(X)

2944 2 A54849
2 1763 2 S16366
44.7 1549 2 I48103
44.4 1690 1 CGHU1B
44.2 1603 2 S23810
43.6 1670 1 CGHU3B
43.6 1670 1 CGHU3B
43.5 920 2 A5748
43.0 1669 1 CGHU4B
43.0 1758 2 T29350
42.7 1669 1 CGMS4B
42.7 1759 2 T29351
42.6 812 2 S31521
42.5 1712 1 CGHU2B
41.9 754 2 A55267
41.7 921 2 S42617
41.6 1142 2 JX0369
41.6

J. Biol. Chem. 260, 2315-2320, 1985
A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter and enhancer regions.
A/Reference number: I55237; MUID:85130970; PMID:2857713
A/Accession: I55237
A/Status: translation not shown; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-34 <CH2>
A/Cross-references: GB:M10627; NID:gl80383; PIDN:AAAS1992.1; PID:9553226
R/Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist, J. Biol. Chem. 265, 6312-6317, 1990
A>Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal propeptide.
A/Reference number: A35233; MUID:90202908; PMID:2318855
A/Accession: A35233
A/Molecule type: protein
A/Residues: 33-52 <WR>
A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved propeptide.
R/Wei, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
A>Title: A base substitution in the exon of a collagen gene causes alternative splicing of the alpha 1(I) chain.
A/Reference number: S09400; MUID:89356643; PMID:2767050
A/Accession: S09400
A/Molecule type: mRNA
A/Residues: 156-183 <WR>
R/Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) chain of collagen.
A/Reference number: A90567; MUID:71038625; PMID:5529814
A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A/Accession: B90567
A/Molecule type: protein
A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 238, 'Z', 241-242, 'Z', 244-245, 'Z', 247-248, 'Z', 251-252, 'Z', 254-255, 'Z', 257-258, 'Z', 261-262, 'Z', 264-265, 'Z', 267-268, 'Z', 271-272, 'Z', 274-275, 'Z', 277-278, 'Z', 281-282, 'Z', 284-285, 'Z', 287-288, 'Z', 291-292, 'Z', 294-295, 'Z', 297-298, 'Z', 301-302, 'Z', 304-305, 'Z', 307-308, 'Z', 311-312, 'Z', 314-315, 'Z', 317-318, 'Z', 321-322, 'Z', 324-325, 'Z', 327-328, 'Z', 331-332, 'Z', 334-335, 'Z', 337-338, 'Z', 341-342, 'Z', 344-345, 'Z', 347-348, 'Z', 351-352, 'Z', 354-355, 'Z', 357-358, 'Z', 361-362, 'Z', 364-365, 'Z', 367-368, 'Z', 371-372, 'Z', 374-375, 'Z', 377-378, 'Z', 381-382, 'Z', 384-385, 'Z', 387-388, 'Z', 391-392, 'Z', 394-395, 'Z', 397-398, 'Z', 401-402, 'Z', 404-405, 'Z', 407-408, 'Z', 411-412, 'Z', 414-415, 'Z', 417-418, 'Z', 421-422, 'Z', 424-425, 'Z', 427-428, 'Z', 431-432, 'Z', 434-435, 'Z', 437-438, 'Z', 441-442, 'Z', 444-445, 'Z', 447-448, 'Z', 451-452, 'Z', 454-455, 'Z', 457-458, 'Z', 461-462, 'Z', 464-465, 'Z', 467-468, 'Z', 471-472, 'Z', 474-475, 'Z', 477-478, 'Z', 481-482, 'Z', 484-485, 'Z', 487-488, 'Z', 491-492, 'Z', 494-495, 'Z', 497-498, 'Z', 501-502, 'Z', 504-505, 'Z', 507-508, 'Z', 511-512, 'Z', 514-515, 'Z', 517-518, 'Z', 521-522, 'Z', 524-525, 'Z', 527-528, 'Z', 531-532, 'Z', 534-535, 'Z', 537-538, 'Z', 541-542, 'Z', 544-545, 'Z', 547-548, 'Z', 551-552, 'Z', 554-555, 'Z', 557-558, 'Z', 561-562, 'Z', 564-565, 'Z', 567-568, 'Z', 571-572, 'Z', 574-575, 'Z', 577-578, 'Z', 581-582, 'Z', 584-585, 'Z', 587-588, 'Z', 591-592, 'Z', 594-595, 'Z', 597-598, 'Z', 601-602, 'Z', 604-605, 'Z', 607-608, 'Z', 611-612, 'Z', 614-615, 'Z', 617-618, 'Z', 621-622, 'Z', 624-625, 'Z', 627-628, 'Z', 631-632, 'Z', 634-635, 'Z', 637-638, 'Z', 641-642, 'Z', 644-645, 'Z', 647-648, 'Z', 651-652, 'Z', 654-655, 'Z', 657-658, 'Z', 661-662, 'Z', 664-665, 'Z', 667-668, 'Z', 671-672, 'Z', 674-675, 'Z', 677-678, 'Z', 681-682, 'Z', 684-685, 'Z', 687-688, 'Z', 691-692, 'Z', 694-695, 'Z', 697-698, 'Z', 701-702, 'Z', 704-705, 'Z', 707-708, 'Z', 711-712, 'Z', 714-715, 'Z', 717-718, 'Z', 721-722, 'Z', 724-725, 'Z', 727-728, 'Z', 731-732, 'Z', 734-735, 'Z', 737-738, 'Z', 741-742, 'Z', 744-745, 'Z', 747-748, 'Z', 751-752, 'Z', 754-755, 'Z', 757-758, 'Z', 761-762, 'Z', 764-765, 'Z', 767-768, 'Z', 771-772, 'Z', 774-775, 'Z', 777-778, 'Z', 781-782, 'Z', 784-785, 'Z', 787-788, 'Z', 791-792, 'Z', 794-795, 'Z', 797-798, 'Z', 801-802, 'Z', 804-805, 'Z', 807-808, 'Z', 811-812, 'Z', 814-815, 'Z', 817-818, 'Z', 821-822, 'Z', 824-825, 'Z', 827-828, 'Z', 831-832, 'Z', 834-835, 'Z', 837-838, 'Z', 841-842, 'Z', 844-845, 'Z', 847-848, 'Z', 851-852, 'Z', 854-855, 'Z', 857-858, 'Z', 861-862, 'Z', 864-865, 'Z', 867-868, 'Z', 871-872, 'Z', 874-875, 'Z', 877-878, 'Z', 881-882, 'Z', 884-885, 'Z', 887-888, 'Z', 891-892, 'Z', 894-895, 'Z', 897-898, 'Z', 901-902, 'Z', 904-905, 'Z', 907-908, 'Z', 911-912, 'Z', 914-915, 'Z', 917-918, 'Z', 921-922, 'Z', 924-925, 'Z', 927-928, 'Z', 931-932, 'Z', 934-935, 'Z', 937-938, 'Z', 941-942, 'Z', 944-945, 'Z', 947-948, 'Z', 951-952, 'Z', 954-955, 'Z', 957-958, 'Z', 961-962, 'Z', 964-965, 'Z', 967-968, 'Z', 971-972, 'Z', 974-975, 'Z', 977-978, 'Z', 981-982, 'Z', 984-985, 'Z', 987-988, 'Z', 991-992, 'Z', 994-995, 'Z', 997-998, 'Z', 1001-1002, 'Z', 1004-1005, 'Z', 1007-1008, 'Z', 1011-1012, 'Z', 1014-1015, 'Z', 1017-1018, 'Z', 1021-1022, 'Z', 1024-1025, 'Z', 1027-1028, 'Z', 1031-1032, 'Z', 1034-1035, 'Z', 1037-1038, 'Z', 1041-1042, 'Z', 1044-1045, 'Z', 1047-1048, 'Z', 1051-1052, 'Z', 1054-1055, 'Z', 1057-1058, 'Z', 1061-1062, 'Z', 1064-1065, 'Z', 1067-1068, 'Z', 1071-1072, 'Z', 1074-1075, 'Z', 1077-1078, 'Z', 1081-1082, 'Z', 1084-1085, 'Z', 1087-1088, 'Z', 1091-1092, 'Z', 1094-1095, 'Z', 1097-1098, 'Z', 1101-1102, 'Z', 1104-1105, 'Z', 1107-1108, 'Z', 1111-1112, 'Z', 1114-1115, 'Z', 1117-1118, 'Z', 1121-1122, 'Z', 1124-1125, 'Z', 1127-1128, 'Z', 1131-1132, 'Z', 1134-1135, 'Z', 1137-1138, 'Z', 1141-1142, 'Z', 1144-1145, 'Z', 1147-1148, 'Z', 1151-1152, 'Z', 1154-1155, 'Z', 1157-1158, 'Z', 1161-1162, 'Z', 1164-1165, 'Z', 1167-1168, 'Z', 1171-1172, 'Z', 1174-1175, 'Z', 1177-1178, 'Z', 1181-1182, 'Z', 1184-1185, 'Z', 1187-1188, 'Z', 1191-1192, 'Z', 1194-1195, 'Z', 1197-1198, 'Z', 1201-1202, 'Z', 1204-1205, 'Z', 1207-1208, 'Z', 1211-1212, 'Z', 1214-1215, 'Z', 1217-1218, 'Z', 1221-1222, 'Z', 1224-1225, 'Z', 1227-1228, 'Z', 1231-1232, 'Z', 1234-1235, 'Z', 1237-1238, 'Z', 1241-1242, 'Z', 1244-1245, 'Z', 1247-1248, 'Z', 1251-1252, 'Z', 1254-1255, 'Z', 1257-1258, 'Z', 1261-1262, 'Z', 1264-1265, 'Z', 1267-1268, 'Z', 1271-1272, 'Z', 1274-1275, 'Z', 1277-1278, 'Z', 1281-1282, 'Z', 1284-1285, 'Z', 1287-1288, 'Z', 1291-1292, 'Z', 1294-1295, 'Z', 1297-1298, 'Z', 1301-1302, 'Z', 1304-1305, 'Z', 1307-1308, 'Z', 1311-1312, 'Z', 1314-1315, 'Z', 1317-1318, 'Z', 1321-1322, 'Z', 1324-1325, 'Z', 1327-1328, 'Z', 1331-1332, 'Z', 1334-1335, 'Z', 1337-1338, 'Z', 1341-1342, 'Z', 1344-1345, 'Z', 1347-1348, 'Z', 1351-1352, 'Z', 1354-1355, 'Z', 1357-1358, 'Z', 1361-1362, 'Z', 1364-1365, 'Z', 1367-1368, 'Z', 1371-1372, 'Z', 1374-1375, 'Z', 1377-1378, 'Z', 1381-1382, 'Z', 1384-1385, 'Z', 1387-1388, 'Z', 1391-1392, 'Z', 1394-1395, 'Z', 1397-1398, 'Z', 1401-1402, 'Z', 1404-1405, 'Z', 1407-1408, 'Z', 1411-1412, 'Z', 1414-1415, 'Z', 1417-1418, 'Z', 1421-1422, 'Z', 1424-1425, 'Z', 1427-1428, 'Z', 1431-1432, 'Z', 1434-1435, 'Z', 1437-1438, 'Z', 1441-1442, 'Z', 1444-1445, 'Z', 1447-1448, 'Z', 1451-1452, 'Z', 1454-1455, 'Z', 1457-1458, 'Z', 1461-1462, 'Z', 1464-1465, 'Z', 1467-1468, 'Z', 1471-1472, 'Z', 1474-1475, 'Z', 1477-1478, 'Z', 1481-1482, 'Z', 1484-1485, 'Z', 1487-1488, 'Z', 1491-1492, 'Z', 1494-1495, 'Z', 1497-1498, 'Z', 1501-1502, 'Z', 1504-1505, 'Z', 1507-1508, 'Z', 1511-1512, 'Z', 1514-1515, 'Z', 1517-1518, 'Z', 1521-1522, 'Z', 1524-1525, 'Z', 1527-1528, 'Z', 1531-1532, 'Z', 1534-1535, 'Z', 1537-1538, 'Z', 1541-1542, 'Z', 1544-1545, 'Z', 1547-1548, 'Z', 1551-1552, 'Z', 1554-1555, 'Z', 1557-1558, 'Z', 1561-1562, 'Z', 1564-1565, 'Z', 1567-1568, 'Z', 1571-1572, 'Z', 1574-1575, 'Z', 1577-1578, 'Z', 1581-1582, 'Z', 1584-1585, 'Z', 1587-1588, 'Z', 1591-1592, 'Z', 1594-1595, 'Z', 1597-1598, 'Z', 1601-1602, 'Z', 1604-1605, 'Z', 1607-1608, 'Z', 1611-1612, 'Z', 1614-1615, 'Z', 1617-1618, 'Z', 1621-1622, 'Z', 1624-1625, 'Z', 1627-1628, 'Z', 1631-1632, 'Z', 1634-1635, 'Z', 1637-1638, 'Z', 1641-1642, 'Z', 1644-1645, 'Z', 1647-1648, 'Z', 1651-1652, 'Z', 1654-1655, 'Z', 1657-1658, 'Z', 1661-1662, 'Z', 1664-1665, 'Z', 1667-1668, 'Z', 1671-1672, 'Z', 1674-1675, 'Z', 1677-1678, 'Z', 1681-1682, 'Z', 1684-1685, 'Z', 1687-1688, 'Z', 1691-1692, 'Z', 1694-1695, 'Z', 1697-1698, 'Z', 1701-1702, 'Z', 1704-1705, 'Z', 1707-1708, 'Z', 1711-1712, 'Z', 1714-1715, 'Z', 1717-1718, 'Z', 1721-1722, 'Z', 1724-1725, 'Z', 1727-1728, 'Z', 1731-1732, 'Z', 1734-1735, 'Z', 1737-1738, 'Z', 1741-1742, 'Z', 1744-1745, 'Z', 1747-1748, 'Z', 1751-1752, 'Z', 1754-1755, 'Z', 1757-1758, 'Z', 1761-1762, 'Z', 1764-1765, 'Z', 1767-1768, 'Z', 1771-1772, 'Z', 1774-1775, 'Z', 1777-1778, 'Z', 1781-1782, 'Z', 1784-1785, 'Z', 1787-1788, 'Z', 1791-1792, 'Z', 1794-1795, 'Z', 1797-1798, 'Z', 1801-1802, 'Z', 1804-1805, 'Z', 1807-1808, 'Z', 1811-1812, 'Z', 1814-1815, 'Z', 1817-1818, 'Z', 1821-1822, 'Z', 1824-1825, 'Z', 1827-1828, 'Z', 1831-1832, 'Z', 1834-1835, 'Z', 1837-1838, 'Z', 1841-1842, 'Z', 1844-1845, 'Z', 1847-1848, 'Z', 1851-1852, 'Z', 1854-1855, 'Z', 1857-1858, 'Z', 1861-1862, 'Z', 1864-1865, 'Z', 1867-1868, 'Z', 1871-1872, 'Z', 1874-1875, 'Z', 1877-1878, 'Z', 1881-1882, 'Z', 1884-1885, 'Z', 1887-1888, 'Z', 1891-1892, 'Z', 1894-1895, 'Z', 1897-1898, 'Z', 1901-1902, 'Z', 1904-1905, 'Z', 1907-1908, 'Z', 1911-1912, 'Z', 1914-1915, 'Z', 1917-1918, 'Z', 1921-1922, 'Z', 1924-1925, 'Z', 1927-1928, 'Z', 1931-1932, 'Z', 1934-1935, 'Z', 1937-1938, 'Z', 1941-1942, 'Z', 1944-1945, 'Z', 1947-1948, 'Z', 1951-1952, 'Z', 1954-1955, 'Z', 1957-1958, 'Z', 1961-1962, 'Z', 1964-1965, 'Z', 1967-1968, 'Z', 1971-1972, 'Z', 1974-1975, 'Z', 1977-1978, 'Z', 1981-1982, 'Z', 1984-1985, 'Z', 1987-1988, 'Z', 1991-1992, 'Z', 1994-1995, 'Z', 1997-1998, 'Z', 2001-2002, 'Z', 2004-2005, 'Z', 2007-2008, 'Z', 2011-2012, 'Z', 2014-2015, 'Z', 2017-2018, 'Z', 2021-2022, 'Z', 2024-2025, 'Z', 2027-2028, 'Z', 2031-2032, 'Z', 2034-2035, 'Z', 2037-2038, 'Z', 2041-2042, 'Z', 2044-2045, 'Z', 2047-2048, 'Z', 2051-2052, 'Z', 2054-2055, 'Z', 2057-2058, 'Z', 2061-2062, 'Z', 2064-2065, 'Z', 2067-2068, 'Z', 2071-2072, 'Z', 2074-2075, 'Z', 2077-2078, 'Z', 2081-2082, 'Z', 2084-2085, 'Z', 2087-2088, 'Z', 2091-2092, 'Z', 2094-2095, 'Z', 2097-2098, 'Z', 2101-2102, 'Z', 2104-2105, 'Z', 2107-2108, 'Z', 2111-2112, 'Z', 2114-2115, 'Z', 2117-2118, 'Z', 2121-2122, 'Z', 2124-2125, 'Z', 2127-2128, 'Z', 2131-2132, 'Z', 2134-2135, 'Z', 2137-2138, 'Z', 2141-2142, 'Z', 2144-2145, 'Z', 2147-2148, 'Z', 2151-2152, 'Z', 2154-2155, 'Z', 2157-2158, 'Z', 2161-2162, 'Z', 2164-2165, 'Z', 2167-2168, 'Z', 2171-2172, 'Z', 2174-2175, 'Z', 2177-2178, 'Z', 2181-2182, 'Z', 2184-2185, 'Z', 2187-2188, 'Z', 2191-2192, 'Z', 2194-2195, 'Z', 2197-2198, 'Z', 2201-2202, 'Z', 2204-2205, 'Z', 2207-2208, 'Z', 2211-2212, 'Z', 2214-2215, 'Z', 2217-2218, 'Z', 2221-2222, 'Z', 2224-2225, 'Z', 2227-2228, 'Z', 2231-2232, 'Z', 2234-2235, 'Z', 2237-2238, 'Z', 2241-2242, 'Z', 2244-2245, 'Z', 2247-2248, 'Z', 2251-2252, 'Z', 2254-2255, 'Z', 2257-2258, 'Z', 2261-2262, 'Z', 2264-2265, 'Z', 2267-2268, 'Z', 2271-2272, 'Z', 2274-2275, 'Z', 2277-2278, 'Z', 2281-2282, 'Z', 2284-2285, 'Z', 2287-2288, 'Z', 2291-2292, 'Z', 2294-2295, 'Z', 2297-2298, 'Z', 2301-2302, 'Z', 2304-2305, 'Z', 2307-2308, 'Z', 2311-2312, 'Z', 2314-2315, 'Z', 2317-2318, 'Z', 2321-2322, 'Z', 2324-2325, 'Z', 2327-2328, 'Z', 2331-2332, 'Z', 2334-2335, 'Z', 2337-2338, 'Z', 2341-2342, 'Z', 2344-2345, 'Z', 2347-2348, 'Z', 2351-2352, 'Z', 2354-2355, 'Z', 2357-2358, 'Z', 2361-2362, 'Z', 2364-2365, 'Z', 2367-2368, 'Z', 2371-2372, 'Z', 2374-2375, 'Z', 2377-2378, 'Z', 2381-2382, 'Z', 2384-2385, 'Z', 2387-2388, 'Z', 2391-2392, 'Z', 2394-2395, 'Z', 2397-2398, 'Z', 2401-2402, 'Z', 2404-2405, 'Z', 2407-2408, 'Z', 2411-2412, 'Z', 2414-2415, 'Z', 2417-2418, 'Z', 2421-2422, 'Z', 2424-2425, 'Z', 2427-2428, 'Z', 2431-2432, 'Z', 2434-2435, 'Z', 2437-2438, 'Z', 2441-2442, 'Z', 2444-2445, 'Z', 2447-2448, 'Z', 2451-2452, 'Z', 2454-2455, 'Z', 2457-2458, 'Z', 2461-2462, 'Z', 2464-2465, 'Z', 2467-2468, 'Z', 2471-2472, 'Z', 2474-2475, 'Z', 2477-2478, 'Z', 2481-2482, 'Z', 2484-2485, 'Z', 2487-2488, 'Z', 2491-2492, 'Z', 2494-2495, 'Z', 2497-2498, 'Z', 2501-2502, 'Z', 2504-2505, 'Z', 2507-2508, 'Z', 2511-2512, 'Z', 2514-2515, 'Z', 2517-2518, 'Z', 2521-2522, 'Z', 2524-2525, 'Z', 2527-2528, 'Z', 2531-2532, 'Z', 2534-2535, 'Z', 2537-2538, 'Z', 2541-2542, 'Z', 2544-2545, 'Z', 2547-2548, 'Z', 2551-2552, 'Z', 2554-2555, 'Z', 2557-2558, 'Z', 2561-2562, 'Z', 2564-2565, 'Z', 2567-2568, 'Z', 2571-2572, 'Z', 2574-2575, 'Z', 2577-2578, 'Z', 2581-2582, 'Z', 2584-2585, 'Z', 2587-2588, 'Z', 2591-2592, 'Z', 2594-2595, 'Z', 2597-2598, 'Z', 2601-2602, 'Z', 2604-2605, 'Z', 2607-2608, 'Z', 2611-2612, 'Z', 2614-2615, 'Z', 2617-2618, 'Z', 2621-2622, 'Z', 2624-2625, 'Z', 2627-2628, 'Z', 2631-2632, 'Z', 2634-2635, 'Z', 2637-2638, 'Z', 2641-2642, 'Z', 2644-2645, 'Z', 2647-2648, 'Z', 2651-2652, 'Z', 2654-2655, 'Z', 2657-2658, 'Z', 2661-2662, 'Z', 2664-2665, 'Z', 2667-2668, 'Z', 2671-2672, 'Z', 2674-2675, 'Z', 2677-2678, 'Z', 2681-2682, 'Z', 2684-2685, 'Z', 2687-2688, 'Z', 2691-2692, 'Z', 2694-2695, 'Z', 2697-2698, 'Z', 2701-2702, 'Z', 2704-2705, 'Z', 2707-2708, 'Z', 2711-2712, 'Z', 2714-2715, 'Z', 2717-2718, 'Z', 2721-2722, 'Z', 2724-2725, 'Z', 2727-2728, 'Z', 2731-2732, 'Z', 2734-2735, 'Z', 2737-2738, 'Z', 2741-2742, 'Z', 2744-2745, 'Z', 2747-2748, 'Z', 2751-2752, 'Z', 2754-2755, 'Z', 2757-2758, 'Z', 2761-2762, 'Z', 2764-2765, 'Z', 2767-2768, 'Z', 2771-2772, 'Z', 2774-2775, 'Z', 2777-2778, 'Z', 2781-2782, 'Z', 2784-2785, 'Z', 2787-2788, 'Z', 2791-2792, 'Z', 2794-2795, 'Z', 2797-2798, 'Z', 2801-2802, 'Z', 2804-2805, 'Z', 2807-2808, 'Z', 2811-2812, 'Z', 2814-2815, 'Z', 2817-2818, 'Z', 2821-2822, 'Z', 2824-2825, 'Z', 2827-2828, 'Z', 2831-2832, 'Z', 2834-2835, 'Z', 2837-2838, 'Z', 2841-2842, 'Z', 2844-2845, 'Z', 2847-2848, 'Z', 2851-2852, 'Z', 2854-2855, 'Z', 2857-2858, 'Z', 2861-2862, 'Z', 2864-2865, 'Z', 2867-2868, 'Z', 2871-2872, 'Z', 2874-2875, 'Z', 2877-2878, 'Z', 2881-2882, 'Z', 2884-2885, 'Z', 2887-2888, 'Z', 2891-2892, 'Z', 2894-2895, 'Z', 2897-2898, 'Z', 2901-2902, 'Z', 2904-2905, 'Z', 2907-2908, 'Z', 2911-2912, 'Z', 2914-2915, 'Z', 2917-2918, 'Z', 2921-2922, 'Z', 2924-2925, 'Z', 2927-2928, 'Z', 2931-2932, 'Z', 2934-2935, 'Z', 2937-2938, 'Z', 2941-2942, 'Z', 2944-2945, 'Z', 2947-2948, 'Z', 2951-2952, 'Z', 2954-2955, 'Z', 2957-2958, 'Z', 2961-2962, 'Z', 2964-2965, 'Z', 2967-2968, 'Z', 2971-2972, 'Z', 2974-2975, 'Z', 2977-2978, 'Z', 2981-2982, 'Z', 2984-2985, 'Z', 2987-2988, 'Z', 2991-2992, 'Z', 2994-2995, 'Z', 2997-2998, 'Z', 3001-3002, 'Z', 3004-3005, 'Z', 3007-3008, 'Z', 3011-3012, 'Z', 3014-3015, 'Z', 3017-3018, 'Z', 3021-3022, 'Z', 3024-3025, 'Z', 3027-3028, 'Z', 3031-3032, 'Z', 3034-3035, 'Z', 3037-3038, 'Z', 3041-3042, 'Z', 3044-3045, 'Z', 3047-3048, 'Z', 3051-3052, 'Z', 3054-3055, 'Z', 3057-3058, 'Z', 3061-3062, 'Z', 3064-3065, 'Z', 3067-3068, 'Z


```
QY 181 GSPGSPDGTGTPGAGDGRPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 240
Db 531 GSPGSPDGTGTPGAGDGRPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 590
QY 241 GPPGAVGPAGKGEAGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 300
Db 591 GPPGAVGPAGKGEAGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 650
QY 301 GVPDGLGAPGSPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 360
Db 651 GVPDGLGAPGSPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 710
QY 361 GAPLEGMPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 420
Db 711 GAPLEGMPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 770
QY 421 GESGSPGAGTGGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 480
Db 771 GESGSPGAGTGGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 830
QY 481 GPAGPAGPPGPIGNDVAGPAGKAGSAGPPGATGFCAGAGVGMFGPPGPPGPA 540
Db 831 GPAGPAGPPGPIGNDVAGPAGKAGSAGPPGATGFCAGAGVGMFGPPGPPGPA 590
QY 541 GKEG 544
Db 891 GKEG 894
```

RESULT 3

CGCHS
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C:Species: Gallus gallus (chicken)
C>Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C:Accession: A90181; A02857
R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A:Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete primary structure of the alpha1(I)-C88 and the complete primary structure of the alpha1(I)-C88
A:Reference number: A90458; MUID:82231995; PMID:7093229
A:Accession: A90458
A:Molecule type: protein
A:Residues: 1-1036 <HIG>
A:Experimental source: skin
A:Note: This is the latest in a series of papers from these workers elucidating the sequence of the alpha1(I) chain of chicken collagen.
R:Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the alpha1(I) chain of chicken collagen.
A:Reference number: A90181; MUID:72243016; PMID:5047697
A:Accession: A90181
A:Molecule type: protein
A:Residues: 1037-1042 <EVR>
A:Experimental source: Skin
A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some species. Most of the prolines at the third position of the tripeptide repeating unit (Pro-1002) is the only 3-hydroxyproline and the only hydroxylated proline in the collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglyutamic acid; trimer; F1/Modified site: Pyroglutamic carboxylic acid (Gln) #status experimental

Query Match 90.4%; Score 2775; DB 1; Length 1042;
Best Local Similarity 89.0%; Pred. NO. 6.9e-128;
Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

```
QY 1 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 60
Db 200 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 259
QY 61 GPGGPPGKDSGPPGAGSGDGTGAKGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 120
Db 260 GPGGPPGKDSGPPGAGSGDGTGAKGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 319
QY 121 GPPGRRGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 180
```

```
Db 320 GPAGRGAPGSRGPPGAGDGTGAGPKGPPGERGSPGAVGPKSPGAGRPGEAGLPGAKGLT 379
QY 181 GSPGSPDGTGTPGAGDGRPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 240
Db 380 GSPGSPDGTGTPGAGDGRPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 439
QY 241 GPPGAVGPAGKGEAGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 300
Db 440 GPPGAVGPAGKGEAGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 499
QY 301 GVPDGLGAPGSPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 360
Db 500 GVPDGLGAPGSPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 559
QY 361 GAPLEGMPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 420
Db 560 GAPLEGMPGARGGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 619
QY 421 GESGSPGAGTGGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 480
Db 620 GESGSPGAGTGGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 679
QY 481 GPAGPAGPPGPIGNDVAGPAGKAGSAGPPGATGFCAGAGVGMFGPPGPPGPA 540
Db 680 GPAGPAGPPGPIGNDVAGPAGKAGSAGPPGATGFCAGAGVGMFGPPGPPGPA 739
QY 541 GKEG 544
Db 740 GKZG 743
```

RESULT 4

T45467
collagen alpha 1(I) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
C:Accession: T45467
R:Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in cells
A:Reference number: Z22977
A:Accession: T45467
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: EMBL:U62528; PIDN:AAB05773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 72.3%; Score 2220; DB 2; Length 1418;
Best Local Similarity 72.6%; Pred. No. 6.5e-101;
Matches 395; Conservative 37; Mismatches 112; Indels 0; Gaps 0;

```
QY 1 GSEGEVGRGPPGPPGARGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 60
Db 315 GPEGAGGPPGPPGARGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 374
QY 61 GPGGPPGKDSGPPGAGSGDGTGAKGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 120
Db 375 GATGPPGKDSGPPGAGSGDGTGAKGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 434
QY 121 GPPGRRGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 180
Db 435 GPPGRRGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 494
QY 181 GSPGSPDGTGTPGAGDGRPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 240
Db 495 GPPGSPDGTGTPGAGDGRPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 554
QY 241 GPPGAVGPAGKGEAGAGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 300
Db 555 GAPGLGPPGKDSGPPGAGSGDGTGAKGPPGPPGARGAGVGMFGPKCAAGRGKAGRGV 614
```

A;Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34663.1.; PID:g30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal propeptide
A;Reference number: A35428; MUID:90285153; PMID:2355003
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81, 'L', 83-103 <RYA2>
A;Note: alternative splice form 2; splicing appears to be under developmental regulation;
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide con-
A;Reference number: A30147; MUID:89233138; PMID:2714801
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157, 'P', 159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GI:
R;Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pr-
A;Reference number: A94227; MUID:90370826; PMID:1975693
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172, 'C', 174-175 <ALA>
A;Note: mutant sequence from a family with family with primary generalized osteoarthritis;
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular
A;Reference number: S64673; MUID:96195147; PMID:8660302
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189, 'X', 191-195; 1224-1230, 'X', 1233-1236 <DIA>
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage, E.;
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil ca
A;Reference number: S63514; MUID:96096730; PMID:8529631
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261, 575-590; 756-763, 'X', 765-779 <FRA>
R;Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G-51VSU20) in the type II collagen gene (COL2A1) in a
A;Reference number: I38867; MUID:95150028; PMID:7847372
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 440, 'G', 442-456, 'B', 458-480, 'P', 482-509 <TIL1>
A;Cross-references: EMBL:U15195; NID:g557053; PIDN:AAB60370.1.; PID:g557054
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>
A;Cross-references: EMBL:X13793; NID:g30037; PIDN:CAA32030.1.; PID:g930050
R;Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
A;Reference number: S05000; MUID:89325561; PMID:2753125
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640, 'A', 642-785 <VIK2>
A;Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1.; PID:g1335018; PIDN:CAA342
PIDN:CAA34283.1.; PID:g1335023; PIDN:CAA34284.1.; PID:g1335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, I
J. Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(I) chain pr-
A;Reference number: A44309; MUID:93054548; PMID:1429602
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA

A/Cross-references: GB:M23759; NID:G180845; EMBL:X03320; GB:M24938; NID:G30100
A/Note: the GenBank PID is based on an incorrect reading frame
A/Accession: I37250
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 541-560
A/Cross-references: EMBL:X02378; GB:M23870; NID:G30107; PIDN:CAA26227.1; PID:G30107
A/Accession: I37251

Query Match	71.7%	Score 2202,	DB 1;	Length 1487;
Best Local Similarity	71.9%;	Pred. No. 5e-100;		
Matches 391;	Conservative 40;	Mismatches 113;	Indels 0;	Gaps 0;
QY	1	GSEGGEGVRGSPGPPGAGAGPAGDPCADGECAGKADGAPCIAGAPGPPGAGRSPGE	60	
Db	384	GPEGAQGRGEGFTGSPGPACASGNPTGICPAKGSAGAPCIAGAPFPFGRGPPDPQ	443	
QY	61	GGCGPPGKGSBPFGAPGSGKDTGAKGEGPCPVGVEGPPGPAEGEKPARGFPPTGLP	120	
Db	444	GATGTLGPKGQTGKPTGAGTKEGQPKGEPQAGPPQAPQAPGAEKKRGARGEPCGVGPI	503	
QY	121	GPBERGGPSRGPPGADGVAGPKPAGERGSPGAPGKSPGEPAGRPGBAGLPGAKGLT	180	
Db	504	GPBERGAPGNRFPQGDLAGPKAPGERGFSGLAGFKGANGDPGRPGEPFLPGARGLT	563	
QY	181	GSPPSGPDGKTPPGPAGEDGRPGPPGPPGARGAEAGVMGFPKPAAGFPGKAGRGV	240	
Db	564	GRPDAGPQKVGPSARGEDGRPGPPGQCARQPCVMGFPKPKGANGFPGKAGEKGLP	623	
QY	241	GPFGVAGPKDGEAGAGCPGPPAGPAGERGEEGPAGSPQFEGLPGPAGPGEAGKPGEE	300	
Db	624	GAPGLRGILFGKDGETGAEGLPPGAPAGERGEQAGPFSQGLPGPPGPPGSGKPGDQ	683	
QY	301	GVPGDLGAPGPSARGEGEPFGERGVEGPPGFPAGPPGADGAPGDGAKGDAGAPAGSE	360	
Db	684	GVPEGAGAPGLVPRGERGFPGERGSPCAQQLQGRGLPGTPTGTDGPKGASGAPGPPGAQ	743	
QY	361	GARGLEGMPEGRAGLSPGPKDCRDAGPKCADGSPCKDGVRGLTGPIGPPGPAGAPGDK	420	
Db	744	GPPLQGMPEGRGAAGIAGPKDRGDVGEKGPESRPKDGGRLTGPIGPPGAGANGEX	803	
QY	421	GSGPSGPAGPTGARGAPCDRGEPPGPAGTAGPPGADGEPGAKGFPBGAGAKGDAGPP	480	
Db	804	GEVPPGPAGSAGARGAPERGETPPTGTSIAGPPGADGQPGAKGQGEAGQKGDAGAP	863	
QY	481	GPAPGAPGPGPLGDVGAQKARGSAGPPCATGFPGAAGRVGPPGSGDAGPPGPPGA	540	
Db	864	GPQPSGAPCPQPTGTVGPKARGAQCPGATGFPGAAGRVGPPGSGNGNPGPPPGPS	923	
QY	541	GKEX 544		
Db	924	GKDG 927		

RESULT 6

A41182
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182; A44885
R:Matsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure and organization.
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: A41182
A:Status: preliminary; not compared with conceptual translation
A:Residues: 1-1419 <MET>
A:Cross-references: GB|M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to the developing embryo.
A:Reference number: A44885; MUID:91347939; PMID:1879363

J. Biol. Chem. 242, 2572-2574, 1967
A>Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A:Reference number: A92029; MUID:67165368; PMID:4290711
A:Contents: CNBx2
A:Accession: A92029
A:Molecule type: Protein
A:Residues: 29-55 <B02>
A:Experimental source: skin and tendon
R:Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A>Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of the hydroxylated peptides of rat skin collagen. The collagenous region of the alpha 1(I) chain of rat skin collagen.
A:Reference number: A90353; MUID:71263178; PMID:4327399
A:Contents: CNBx4
A:Accession: A90353
A:Molecule type: protein
A:Residues: 56-102 <B01>
A:Experimental source: skin
R:Butler, W.T.
Biochemistry 9, 44-50, 1970
A>Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The collagenous region of the alpha 1(I) chain of rat skin collagen.
A:Reference number: A90566; MUID:70085124; PMID:5411206
A:Contents: CNBx5
A:Accession: A90566
A:Molecule type: protein
A:Residues: 103-139 <B02>
A:Experimental source: skin
R:Ballian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A>Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylated peptides of rat skin collagen.
A:Reference number: A90357; MUID:72136131; PMID:4335087
A:Contents: CNBx8
A:Accession: A90357
A:Molecule type: protein
A:Residues: 140-238 <B01>
A:Experimental source: skin
R:Ballian, G.; Click, E.M.; Hermanson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A>Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylated peptides of rat skin collagen.
A:Reference number: A90362; MUID:73006942; PMID:4342027
A:Contents: CNBx8
A:Accession: A90362
A:Molecule type: protein
A:Residues: 239-418 <B02>
A:Experimental source: skin
R:Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A>Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of the hydroxylated peptides of rat skin collagen.
A:Reference number: A90379; MUID:7421984; PMID:4366532
A:Contents: CNBx3
A:Accession: A90379
A:Molecule type: protein
A:Residues: 419-567 <B03>
A:Experimental source: skin
R:Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A>Title: Structural and immunogenic properties of a major antigenic determinant in neutral salt-extracted rat skin collagen.
A:Reference number: A91209; MUID:7401954; PMID:4126850
A:Contents: CNBx6
A:Accession: A91209
A:Molecule type: protein
A:Residues: 568-651 <ST1>
A:Experimental source: skin
R:Stoltz, M.; Timpl, R.; Kuehn, K.
A>Note: This region probably corresponds to positions 949-1032 of the alpha 1(I) chain of rat skin collagen.
A:Title: Non-helical regions in rat collagen alpha1-chain.
A:Reference number: A91385; MUID:73049495; PMID:4636751
A:Contents: CNBx6
A:Accession: A91385
A:Molecule type: protein
A:Residues: 651-671 <ST2>
A:Experimental source: skin

A>Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequence of this region (residues 651-671 above) probably corresponds to positions 1032-1052 of the alpha 1(I) chain of rat skin collagen.
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit and subsequently O-glycosylated.
C:Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin collagen is as follows: 1-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99, 100-109, 110-119, 120-129, 130-139, 140-149, 150-159, 160-169, 170-179, 180-189, 190-199, 200-209, 210-219, 220-229, 230-239, 240-249, 250-259, 260-269, 270-279, 280-289, 290-299, 300-309, 310-319, 320-329, 330-339, 340-349, 350-359, 360-369, 370-379, 380-389, 390-399, 400-409, 410-419, 420-429, 430-439, 440-449, 450-459, 460-469, 470-479, 480-489, 490-499, 500-509, 510-519, 520-529, 530-539, 540-549, 550-559, 560-569, 570-579, 580-589, 590-599, 600-609, 610-619, 620-629, 630-639, 640-649, 650-659, 660-669, 670-679, 680-689, 690-699, 700-709, 710-719, 720-729, 730-739, 740-749, 750-759, 760-769, 770-779, 780-789, 790-799, 800-809, 810-819, 820-829, 830-839, 840-849, 850-859, 860-869, 870-879, 880-889, 890-899, 900-909, 910-919, 920-929, 930-939, 940-949, 950-959, 960-969, 970-979, 980-989, 990-999, 1000-1009, 1010-1019, 1020-1029, 1030-1039, 1040-1049, 1050-1059, 1060-1069, 1070-1079, 1080-1089, 1090-1099, 1100-1109, 1110-1119, 1120-1129, 1130-1139, 1140-1149, 1150-1159, 1160-1169, 1170-1179, 1180-1189, 1190-1199, 1200-1209, 1210-1219, 1220-1229, 1230-1239, 1240-1249, 1250-1259, 1260-1269, 1270-1279, 1280-1289, 1290-1299, 1300-1309, 1310-1319, 1320-1329, 1330-1339, 1340-1349, 1350-1359, 1360-1369, 1370-1379, 1380-1389, 1390-1399, 1400-1409, 1410-1419, 1420-1429, 1430-1439, 1440-1449, 1450-1459, 1460-1469, 1470-1479, 1480-1489, 1490-1499, 1500-1509, 1510-1519, 1520-1529, 1530-1539, 1540-1549, 1550-1559, 1560-1569, 1570-1579, 1580-1589, 1590-1599, 1600-1609, 1610-1619, 1620-1629, 1630-1639, 1640-1649, 1650-1659, 1660-1669, 1670-1679, 1680-1689, 1690-1699, 1700-1709, 1710-1719, 1720-1729, 1730-1739, 1740-1749, 1750-1759, 1760-1769, 1770-1779, 1780-1789, 1790-1799, 1800-1809, 1810-1819, 1820-1829, 1830-1839, 1840-1849, 1850-1859, 1860-1869, 1870-1879, 1880-1889, 1890-1899, 1900-1909, 1910-1919, 1920-1929, 1930-1939, 1940-1949, 1950-1959, 1960-1969, 1970-1979, 1980-1989, 1990-1999, 2000-2009, 2010-2019, 2020-2029, 2030-2039, 2040-2049, 2050-2059, 2060-2069, 2070-2079, 2080-2089, 2090-2099, 2100-2109, 2110-2119, 2120-2129, 2130-2139, 2140-2149, 2150-2159, 2160-2169, 2170-2179, 2180-2189, 2190-2199, 2200-2209, 2210-2219, 2220-2229, 2230-2239, 2240-2249, 2250-2259, 2260-2269, 2270-2279, 2280-2289, 2290-2299, 2300-2309, 2310-2319, 2320-2329, 2330-2339, 2340-2349, 2350-2359, 2360-2369, 2370-2379, 2380-2389, 2390-2399, 2400-2409, 2410-2419, 2420-2429, 2430-2439, 2440-2449, 2450-2459, 2460-2469, 2470-2479, 2480-2489, 2490-2499, 2500-2509, 2510-2519, 2520-2529, 2530-2539, 2540-2549, 2550-2559, 2560-2569, 2570-2579, 2580-2589, 2590-2599, 2600-2609, 2610-2619, 2620-2629, 2630-2639, 2640-2649, 2650-2659, 2660-2669, 2670-2679, 2680-2689, 2690-2699, 2700-2709, 2710-2719, 2720-2729, 2730-2739, 2740-2749, 2750-2759, 2760-2769, 2770-2779, 2780-2789, 2790-2799, 2800-2809, 2810-2819, 2820-2829, 2830-2839, 2840-2849, 2850-2859, 2860-2869, 2870-2879, 2880-2889, 2890-2899, 2900-2909, 2910-2919, 2920-2929, 2930-2939, 2940-2949, 2950-2959, 2960-2969, 2970-2979, 2980-2989, 2990-2999, 3000-3009, 3010-3019, 3020-3029, 3030-3039, 3040-3049, 3050-3059, 3060-3069, 3070-3079, 3080-3089, 3090-3099, 3100-3109, 3110-3119, 3120-3129, 3130-3139, 3140-3149, 3150-3159, 3160-3169, 3170-3179, 3180-3189, 3190-3199, 3200-3209, 3210-3219, 3220-3229, 3230-3239, 3240-3249, 3250-3259, 3260-3269, 3270-3279, 3280-3289, 3290-3299, 3300-3309, 3310-3319, 3320-3329, 3330-3339, 3340-3349, 3350-3359, 3360-3369, 3370-3379, 3380-3389, 3390-3399, 3400-3409, 3410-3419, 3420-3429, 3430-3439, 3440-3449, 3450-3459, 3460-3469, 3470-3479, 3480-3489, 3490-3499, 3500-3509, 3510-3519, 3520-3529, 3530-3539, 3540-3549, 3550-3559, 3560-3569, 3570-3579, 3580-3589, 3590-3599, 3600-3609, 3610-3619, 3620-3629, 3630-3639, 3640-3649, 3650-3659, 3660-3669, 3670-3679, 3680-3689, 3690-3699, 3700-3709, 3710-3719, 3720-3729, 3730-3739, 3740-3749, 3750-3759, 3760-3769, 3770-3779, 3780-3789, 3790-3799, 3800-3809, 3810-3819, 3820-3829, 3830-3839, 3840-3849, 3850-3859, 3860-3869, 3870-3879, 3880-3889, 3890-3899, 3900-3909, 3910-3919, 3920-3929, 3930-3939, 3940-3949, 3950-3959, 3960-3969, 3970-3979, 3980-3989, 3990-3999, 4000-4009, 4010-4019, 4020-4029, 4030-4039, 4040-4049, 4050-4059, 4060-4069, 4070-4079, 4080-4089, 4090-4099, 4100-4109, 4110-4119, 4120-4129, 4130-4139, 4140-4149, 4150-4159, 4160-4169, 4170-4179, 4180-4189, 4190-4199, 4200-4209, 4210-4219, 4220-4229, 4230-4239, 4240-4249, 4250-4259, 4260-4269, 4270-4279, 4280-4289, 4290-4299, 4300-4309, 4310-4319, 4320-4329, 4330-4339, 4340-4349, 4350-4359, 4360-4369, 4370-4379, 4380-4389, 4390-4399, 4400-4409, 4410-4419, 4420-4429, 4430-4439, 4440-4449, 4450-4459, 4460-4469, 4470-4479, 4480-4489, 4490-4499, 4500-4509, 4510-4519, 4520-4529, 4530-4539, 4540-4549, 4550-4559, 4560-4569, 4570-4579, 4580-4589, 4590-4599, 4600-4609, 4610-4619, 4620-4629, 4630-4639, 4640-4649, 4650-4659, 4660-4669, 4670-4679, 4680-4689, 4690-4699, 4700-4709, 4710-4719, 4720-4729, 4730-4739, 4740-4749, 4750-4759, 4760-4769, 4770-4779, 4780-4789, 4790-4799, 4800-4809, 4810-4819, 4820-4829, 4830-4839, 4840-4849, 4850-4859, 4860-4869, 4870-4879, 4880-4889, 4890-4899, 4900-4909, 4910-4919, 4920-4929, 4930-4939, 4940-4949, 4950-4959, 4960-4969, 4970-4979, 4980-4989, 4990-4999, 5000-5009, 5010-5019, 5020-5029, 5030-5039, 5040-5049, 5050-5059, 5060-5069, 5070-5079, 5080-5089, 5090-5099, 5100-5109, 5110-5119, 5120-5129, 5130-5139, 5140-5149, 5150-5159, 5160-5169, 5170-5179, 5180-5189, 5190-5199, 5200-5209, 5210-5219, 5220-5229, 5230-5239, 5240-5249, 5250-5259, 5260-5269, 5270-5279, 5280-5289, 5290-5299, 5300-5309, 5310-5319, 5320-5329, 5330-5339, 5340-5349, 5350-5359, 5360-5369, 5370-5379, 5380-5389, 5390-5399, 5400-5409, 5410-5419, 5420-5429, 5430-5439, 5440-5449, 5450-5459, 5460-5469, 5470-5479, 5480-5489, 5490-5499, 5500-5509, 5510-5519, 5520-5529, 5530-5539, 5540-5549, 5550-5559, 5560-5569, 5570-5579, 5580-5589, 5590-5599, 5600-5609, 5610-5619, 5620-5629, 5630-5639, 5640-5649, 5650-5659, 5660-5669, 5670-5679, 5680-5689, 5690-5699, 5700-5709, 5710-5719, 5720-5729, 5730-5739, 5740-5749, 5750-5759, 5760-5769, 5770-5779, 5780-5789, 5790-5799, 5800-5809, 5810-5819, 5820-5829, 5830-5839, 5840-5849, 5850-5859, 5860-5869, 5870-5879, 5880-5889, 5890-5899, 5900-5909, 5910-5919, 5920-5929, 5930-5939, 5940-5949, 5950-5959, 5960-5969, 5970-5979, 5980-5989, 5990-5999, 6000-6009, 6010-6019, 6020-6029, 6030-6039, 6040-6049, 6050-6059, 6060-6069, 6070-6079, 6080-6089, 6090-6099, 6100-6109, 6110-6119, 6120-6129, 6130-6139, 6140-6149, 6150-6159, 6160-6169, 6170-6179, 6180-6189, 6190-6199, 6200-6209, 6210-6219, 6220-6229, 6230-6239, 6240-6249, 6250-6259, 6260-6269, 6270-6279, 6280-6289, 6290-6299, 6300-6309, 6310-6319, 6320-6329, 6330-6339, 6340-6349, 6350-6359, 6360-6369, 6370-6379, 6380-6389, 6390-6399, 6400-6409, 6410-6419, 6420-6429, 6430-6439, 6440-6449, 6450-6459, 6460-6469, 6470-6479, 6480-6489, 6490-6499, 6500-6509, 6510-6519, 6520-6529, 6530-6539, 6540-6549, 6550-6559, 6560-6569, 6570-6579, 6580-6589, 6590-6599, 6600-6609, 6610-6619, 6620-6629, 6630-6639, 6640-6649, 6650-6659, 6660-6669, 6670-6679, 6680-6689, 6690-6699, 6700-6709, 6710-6719, 6720-6729, 6730-6739, 6740-6749, 6750-6759, 6760-6769, 6770-6779, 6780-6789, 6790-6799, 6800-6809, 6810-6819, 6820-6829, 6830-6839, 6840-6849, 6850-6859, 6860-6869, 6870-6879, 6880-6889, 6890-6899, 6900-6909, 6910-6919, 6920-6929, 6930-6939, 6940-6949, 6950-6959, 6960-6969, 6970-6979, 6980-6989, 6990-6999, 7000-7009, 7010-7019, 7020-7029, 7030-7039, 7040-7049, 7050-7059, 7060-7069, 7070-7079, 7080-7089, 7090-7099, 7100-7109, 7110-7119, 7120-7129, 7130-7139, 7140-7149, 7150-7159, 7160-7169, 7170-7179, 7180-7189, 7190-7199, 7200-7209, 7210-7219, 7220-7229, 7230-7239, 7240-7249, 7250-7259, 7260-7269, 7270-7279, 7280-7289, 7290-7299, 7300-7309, 7310-7319, 7320-7329, 7330-7339, 7340-7349, 7350-7359, 7360-7369, 7370-7379, 7380-7389, 7390-7399, 7400-7409, 7410-7419, 7420-7429, 7430-7439, 7440-7449, 7450-7459, 7460-7469, 7470-7479, 7480-7489, 7490-7499, 7500-7509, 7510-7519, 7520-7529, 7530-7539, 7540-7549, 7550-7559, 7560-7569, 7570-7579, 7580-7589, 7590-7599, 7600-7609, 7610-7619, 7620-7629, 7630-7639, 7640-7649, 7650-7659, 7660-7669, 7670-7679, 7680-7689, 7690-7699, 7700-7709, 7710-7719, 7720-7729, 7730-7739, 7740-7749, 7750-7759, 7760-7769, 7770-7779, 7780-7789, 7790-7799, 7800-7809, 7810-7819, 7820-7829, 7830-7839, 7840-7849, 7850-7859, 7860-7869, 7870-7879, 7880-7889, 7890-7899, 7900-7909, 7910-7919, 7920-7929, 7930-7939, 7940-7949, 7950-7959, 7960-7969, 7970-7979, 7980-7989, 7990-7999, 8000-8009, 8010-8019, 8020-8029, 8030-8039, 8040-8049, 8050-8059, 8060-8069, 8070-8079, 8080-8089, 8090-8099, 8100-8109, 8110-8119, 8120-8129, 8130-8139, 8140-8149, 8150-8159, 8160-8169, 8170-8179, 8180-8189, 8190-8199, 8200-8209, 8210-8219, 8220-8229, 8230-8239, 8240-8249, 8250-8259, 8260-8269, 8270-8279, 8280-8289, 8290-8299, 8300-8309, 8310-8319, 8320-8329, 8330-8339, 8340-8349, 8350-8359, 8360-8369, 8370-8379, 8380-8389, 8390-8399, 8400-8409, 8410-8419, 8420-8429, 8430-8439, 8440-8449, 8450-8459, 8460-8469, 8470-8479, 8480-8489, 8490-8499, 8500-8509, 8510-8519, 8520-8529, 8530-8539, 8540-8549, 8550-8559, 8560-8569, 8570-8579, 8580-8589, 8590-8599, 8600-8609, 8610-8619, 8620-8629, 8630-8639, 8640-8649, 8650-8659, 8660-8669, 8670-8679, 8680-8689, 8690-8699, 8700-8709, 8710-8719, 8720-8729, 8730-8739, 8740-8749, 8750-8759, 8760-8769, 8770-8779, 8780-8789, 8790-8799, 8800-8809, 8810-8819, 8820-8829, 8830-8839, 8840-8849, 8850-8859, 8860-8869, 8870-8879, 8880-8889, 8890-8899, 8900-8909, 8910-8919, 8920-8929, 8930-8939, 8940-8949, 8950-8959, 8960-8969, 8970-8979, 8980-8989, 8990-8999, 9000-9009, 9010-9019, 9020-9029, 9030-9039, 9040-9049, 9050-9059, 9060-9069, 9070-9079, 9080-9089, 9090-9099, 9100-9109, 9110-9119, 9120-9129, 9130-9139, 9140-9149, 9150-9159, 9160-9169, 9170-9179, 9180-9189, 9190-9199, 9200-9209, 9210-9219, 9220-9229, 9230-9239, 9240-9249, 9250-9259, 9260-9269, 9270-9279, 9280-9289, 9290-9299, 9300-9309, 9310-9319, 9320-9329, 9330-9339, 9340-9349, 9350-9359, 9360-9369, 9370-9379, 9380-9389, 9390-9399, 9400-9409, 9410-9419, 9420-9429, 9430-9439, 9440-9449, 9450-9459, 9460-9469, 9470-9479, 9480-9489, 9490-9499, 9500-9509, 9510-9519, 9520-9529, 9530-9539, 9540-9549, 9550-9559, 9560-9569, 9570-9579, 9580-9589, 9590-9599, 9600-9609, 9610-9619, 9620-9629, 9630-9639, 9640-9649, 9650-9659, 9660-9669, 9670-9679, 9680-9689, 9690-9699, 9700-9709, 9710-9719, 9720-9729, 9730-9739, 9740-9749, 9750-9759, 9760-9769, 9770-9779, 9780-9789, 9790-9799, 9800-9809, 9810-9819, 9820-9829, 9830-9839, 9840-

F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 69.9%; Score 2147; DB 1; Length 1486;
Best Local Similarity 70.4%; Pred. NO. 2.3e-97;
Matches 383; Conservative 41; Mismatches 120; Indels 0; Gaps 0;

Qy	1	CSEGGVGRBFP	PPGAGAGPAGD	PCADGEC	CAKAGDAG	ACIAGAP	CFPG	CA	GPCSP	60	
Db	386	GPEGAGQGR	BESGTFG	SGFPG	PASGN	PTGTG	IPG	AKSSGG	PGIAG	PFPGRP	445
Qy	61	PGGPGPGK	GDSPG	APGSG	KGDTCA	KGECP	QVGV	GGPP	QGPAG	CEBKPG	120
Db	446	GATGLPK	GETGDP	GVAG	FKGEQ	PKGEIG	SGAG	PG	QAG	PG	505
Qy	121	GPBGRG	QPGSR	PTPGAD	GVAG	PKPAG	ERGSG	FGP	PAKGS	PCEACR	180
Db	506	GPBGRG	APGNR	PGQD	GLAG	PKGAP	GERGV	PG	LGCG	PKGNGD	565
Qy	181	QSPGSP	PGDGTG	PPG	PAGE	DCR	PGP	PP	PCAK	CEAG	240
Db	566	GRPDAG	PGQKV	PSGAS	GED	GRFG	PG	QAG	QGP	VGWGF	625
Qy	241	GPBGAV	PGAKD	GAGAB	PPG	PPAG	PAG	AGER	GGP	AGSG	300
Db	626	GAPGLR	GLPGK	DETGT	SGPN	PZAP	AGER	GEQ	PPG	SGFQ	685
Qy	301	GVPGDL	GAPSG	ARGER	GGP	FGERG	VGPP	PG	PAG	PGD	360
Db	686	GVPG	BAGAP	GLVGR	PRGER	GF	PER	SGSG	PGQL	QGP	745
Qy	361	GAPGLE	MPGR	GAG	LAGL	PGPK	GD	RDA	PKX	ADGS	420
Db	746	GPGLQ	GNM	PGER	GAGT	SGPK	GD	RDT	CEK	GPEG	805
Qy	421	GESGSG	PAGPT	GARG	AGD	RGER	GGP	PG	PAG	PG	480
Db	806	GESGSG	PGPIV	GA	KAP	ED	KG	EN	PP	PPAG	865
Qy	481	GPA	GAP	PGP	FTG	DV	GA	PK	AG	SG	540
Db	866	GP	QPS	GAP	PG	PGT	VG	PG	PK	AG	925

RESULT 10

A40333
Collagen alpha 1(I) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C:Accession: A40333
R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during *Xenopus laevis* em
A:Reference number: A40333; MUID:9201898; PMID:1918153
A:Accession: A40333
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1492 <SUA>
A:Cross-references: GB:M63596
A:Note: This sequence is presented as substitutions relative to another sequence in a fi
es they replace; the appropriate interpretation of the sequence figure was reconstructed
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 69.4%; Score 2131; DB 2; Length 1492;
Best Local Similarity 70.0%; Pred.No. 1.4e-96;
Matches 381; Conservative 41; Mismatches 122; Indels 0; Gaps 0;

Query Match \uparrow 69.4%; Score 2131; DB 2; Length 1492;
Best Local Similarity 70.0%; Pred. NO. 1.4e-96;
Matches 381; Conservative 41; Mismatches 122; Indels 0; Gaps 0;

QY 1 GSEPGVGVGPGPPGAGAACPADPGADGEPGAKGADGAGIAGAPGPGAGSGGE 60
Db 389 GPEGAQGRGEGSGTPGSPGSGASGNPGTDGPGAKGSSGAGIAGAPGPGRGPFGQ 448
QY 61 GPQGGPPGKDGSGEPGAPGSKDGTAKAGPFGFVGVGEGPPGPGAGEEKGPGARGGPGTGLP 120
Db 449 GATGPIGPKGETGDPGLAGFKSEHGPKGILGSAGPQGAFGPAGEEKGARGGEPGAAGPL 508
QY 121 GPPGEGGPGSGRFFPGADGVAGPKGAPAGRGSGPPAGPKGSPGEGAGRGPGEAGLPGAKGLT 180
Db 509 GPPGERGAPGNRGFPQDGLAGPKGAPGGRGVFGLGPKGKNGDPRGPGEPLFGARGLT 568
QY 181 GSPGSGPPDGKTPGPGAGEDGRPGPPGPPGARGGAGVGMGPPGKGAAGEBPKGAGRGVP 240
Db 569 GRPDDAGPQKVGPSAAGEDGRPGPPGQAGKQPGVMGFPQPKGANSEPGKAGEKGLG 628
QY 241 GPPGAVGPPAGKGEAGAEPPGPPGAGPAGRGSEPPAGSPGFEGLPGPAGPPGAGPKGEE 300
Db 629 GAFGLRGLPKDGETGAQCGPNGPAGPAGERGEQGPSPGFGQLGPPGPGSPGSGKPGDQ 688
QY 301 GVPDGLGAPGSCARGEPFPGRGVGGPPGAPGADGAPGDDGAKDGAGAPGPGSE 360
Db 689 GVPEGAAGAGLVPRGERFPFGRSGSGQGLGSGRLGFTPTDGPKGATGSPGNGAQ 748
QY 361 GAPLEGMPCERGAAGLPGPKDGRGADGPKGADGSPGKDGVRGLTGPFGPPGAPAGPDK 420
Db 749 GPFGLQMPGERCAAGISGPKGDRGDTGEGKPGSAGFGKDSRGLTGFLGPPGAPGNGEK 808
QY 421 GESGSGGAPGPTGARGAPDGRGEPGPPGAPGAPGADGEPGAKGSPGDGAKGAGPAP 480
Db 809 GESGSGGPPPIVGARGAPDGRGSGNGPFGPAGFPAGPFGSDGQAGLKGDCQSGKGAGAP 868
QY 481 GPAGPAGPPGPIGDVCAKAGKAGGAGPAGPAGTFPGAAAGRVGPPGPGDAGPPGPGPA 540
Db 869 GPQPGSAGPQPGTGVNGPKGARGAGGAGATGFFGAGAGRVCTPGNGNPPGPPGSA 928
QY 541 KKEG 544
Db 929 KKEG 932

RESULT 11
CGB01S
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence revision 31-Dec-1993 #text change 31-Mar-2000
C:Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048; A02853
R:Rautenberg, J.; Timpl, R.; Furtmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A:Title: Structural characterization of N-terminal antigenic determinants in calf and hu
A:Reference number: A91193; MUID:72255334; PMID:4115172
A:Accession: A91193
A:Molecule type: protein
A:Residues: 1-19 <RAU>
A:Experimental source: skin
A:Note: the epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain, is conv
R:Fietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A:Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromid
A:Reference number: A91229; MUID:76022320; PMID:1164916
A:Accession: A91229
A:Molecule type: protein
A:Residues: 20-145 <FIE>
A:Experimental source: skin
A:Note: Lys-103 is hydroxylated and binds glucosylgalactose
R:Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A:Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3 from cal
A:Reference number: A91387; MUID:73049499; PMID:4673951
A:Accession: A91387
A:Molecule type: protein
A:Residues: 146-294 <FIE>
A:Experimental source: skin

R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 396-400, 1973
A;Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fr
A;Reference number: A91211; MUID:74086118; PMID:4359390
A;Accession: A91211
A;Molecule type: Protein
A;Residues: 295-562 <F13>
A;Experimental source: skin
R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972
A;Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
A;Reference number: A91201; MUID:73042276; PMID:4343808
A;Accession: A91201
A;Molecule type: protein
A;Residues: 563-675 <WEN>
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A;Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C
A;Reference number: A91200; MUID:73042275; PMID:4343807
A;Accession: A91200
A;Molecule type: protein
A;Residues: 676-758 <F14>
A;Experimental source: skin
R;Rautenberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
FEBS Lett 21, 75-79, 1972
A;Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of
A;Reference number: A43048
A;Accession: A43048
A;Molecule type: protein
A;Residues: 759-779 <RA2>
A;Experimental source: skin
C;Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 558, and 670 may be hydrox
C;Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are h
C;Comment: The order of the eight CNBR peptides in the alpha 1(I) chain of bovine skin c
9, 149, 268, and 217 residues.
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 67.8%; Score 2076.5; DB 1; Length 779;
Best Local Similarity 78.0%; Pred. No. 3.7e-94;
Matches 382; Conservative 18; Mismatches 51; Indels 39; Gaps 6;

Qy 55 GPSGPGPGPKGDSGPGAPGSGKGTGAKGPGVGVGPGPGAGGPGKARGBP 114
Db 17 GPMGPGPRGLPGF-----PGA-----POPGFQGP--GEPEPGNSGPM 55

Qy 115 GPTGLPFGPGGPGSRGPGFPGADGVAGPKPAGRGSPGAPGKSGPGAGRPGAGLP 174
Db 56 GPRGPGPGPKNGD-----DGEAGKPRGPRGPRGPGQGAR-----GLPGTAGLP 100

Qy 175 GAKGLTSPSGPDGKTPGPGAGEDGRPGPGPGPGAGGAGVMGFPKGAAGPGKA 234
Db 101 GMKGRGFTGLDGAAGDAGPAGPKGPGSPGNGAPQMGPRGLPGFPGKGAAGEPGKA 160

Qy 235 GERGVPGPPGAVGAGKDGAGAGPPGPGAPGAGRGSGPGAGSGPGFGLPGPAGPGGA 294
Db 161 GERGVPGPPGAVGAGKDGAGAGPPGPGAPGAGRGSGPGAGSGPGFGLPGPAGPGGA 220

Qy 295 KGSGEVPGLGAPGPGSGRGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 354
Db 221 KGPGEQVPGDLGAPGPGSGRGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 280

Qy 355 GAPGSEAGLGEPGGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPFGPGPA 414
Db 281 GAPGSGQAPGLQGPFGERGAAGLPGPKGDRGDAGPKGADGAPKDGVRGLTGPFGPGPA 340

Qy 415 GAPDCKGSPSGPAGTGTARGAPGDRGPGPGPGPAGPAGPAGPAGPAGPAGPAG 474
Db 341 GAPDCKGAGPGSGPA---GTRGAPGDRGPGPGPAGPAGPAGPAGPAGPAGPAG 397

Qy 475 GDAGPPGPGAGPAGPPGPIGVGAPGARGSGAGPPGATGFFGAAAGRVGPPGSGDAGPP 534
Db 398 GDAGPPGPGAGPAGPPGPIGVGAPGARGSGAGPPGATGFFGAAAGRVGPPGSGNAGPP 457

Qy 535 GPGPGAGKEG 544
Db 458 GPGPGAGKEG 467

RESULT 12

CGHU2V
C;Species: Homo sapiens (man)
C;Date: 31-Jul-1989 #sequence_revision 28-Jul-1995 #text_change 31-Dec-2000
C;Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A;Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the stru
A;Reference number: A31427; MUID:89123368; PMID:2914927
A;Accession: A31427
A;Molecule type: mRNA
A;Residues: 1-463 <WOO>
A;Cross-references: GB:J04478; NID:G179697; PIDN:AA51859.1; PID:G179698
A;Experimental source: Placenta
R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A;Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for
A;Reference number: A54555; MUID:92314691; PMID:1820205
A;Accession: A54555
A;Molecule type: DNA
A;Residues: 1-32 <GRE>
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Boutillon, M.M.; Be
Eur. J. Biochem. 221, 987-995, 1994
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.
A;Reference number: S43642; MUID:94237164; PMID:8181482
A;Accession: S43643
A;Molecule type: protein
A;Residues: 288-291, 'P', '293-294, 'X', '296-297, '606, 'X', '608-617 <MOR>
R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A;Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibril
A;Reference number: A25874; MUID:87146331; PMID:3029669
A;Accession: A25874
A;Molecule type: mRNA; DNA
A;Residues: 398-1496 <WEI>
A;Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:G1340175
A;Experimental source: rhabdomyosarcoma cell line
R;Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985
A;Title: Partial covalent structure of the human alpha 2 type V collagen chain.
A;Reference number: I55239; MUID:85182703; PMID:2985598
A;Accession: I55239
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1002-1226 <RE2>
A;Cross-references: GB:M10956; NID:G180427; PIDN:AA52007.1; PID:G180428
A;Note: part of this sequence were determined by protein sequencing
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long ar
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1003-1034 <RES>
A;Cross-references: GB:M11135; NID:G179693; PIDN:AA51857.1; PID:G179694
A;Note: part of this sequence were determined by protein sequencing
R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
J. Biol. Chem. 260, 11216-11222, 1985
A;Title: Complete primary structure of the human alpha-2 type V procollagen COOH-termin
A;Reference number: A25374; MUID:85289337; PMID:2411731

QY	121	GPPRGCGPSEGFADGVAGPKGABGRGSPGPAGPKGSPEGACRPGEAGLPGAKGLT	180
DB	513	GVVSGRGAPGNKGFPSDGLPFKGAQGBRGVPVSSGPKKSQSDPRFPBEPGLPGARGLT	572
QY	181	GSFGSPGDGKTGTPPGAGEDCRPPGPPGGARGBAGVMVFPGPKAAAGEPGKAGRGVP	240
DB	573	GNFVGQPEGKLPLCAPGEDCRPPGPSIGIKGQPDTMGLPGPKGSNGDPGKPGEAGNP	632
QY	241	GGPVAVGPAKGDEAGAEPPCPAGPAGRGEEGPAGSPGFEGLPGPAGPPGEGAKPGEE	300
DB	633	GVFGQRGAPGDKGVGYPGPPGPPGRLRGREGQQPPGTGTGQHHPGPPPTPVEGGHFGDQ	692
QY	301	GVFGDLGAPSPSARGEPGPPGERGVGEGPPGAPGPDGACPDGCAKDAGAPGAPGSE	360
DB	693	GVFGPGFVAVGLPRGCRCNPGERGEPGITGLPGEKMGAGCHGPDGPKGSPGSGTPGDT	752
QY	361	GAPCLEMPBERGNACLGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK	420
DB	753	GGPGQLGMPPGERGIAGTFPKGDRGIGSKGAETAGNDGAGGLDPLGPPGAPGLLGEX	812
QY	421	GSGPSGPGAPTCARCAPCDRGEPPGPPGAPGAPGADCEPGAKEGPDGAGAKGDAGPP	480
DB	813	GEFGPRLVLPVPPCSNRNPGSRGENGTGAVGFAGPQGSDQPGVKGEPEGPFQKGDAGSP	872
QY	481	GPAGPAGPPGPTGDVCAPAKGARGSAGPPCATGFFPMAGRVCPFPSPGCDAGPPGPPGA	540
DB	873	GGOGLAGSPGHGPNVPLKKGRGTGQGPATGTFPSAGRVPFPFGAGAPPAGPLJEP	932
QY	541	GKEG 544	
DB	933	GKEG 936	

RESULT 13
A43291
collagen alpha 2(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A43291; A54328
R:Phillips, C.I.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.
Genomics 13, 1345-1346, 1992
A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen
A:Reference number: A43291; MUID:92372043; PMID:1505972
A:Accession: A43291
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1373 <PHI>
A:Cross-references: GB:X58251; NID:G50489; PIDN:CAA41205.1; PID:G50489
A>Note: sequence extracted from NCBI backbone (NCBIP:112027)
R:Phillips, C.I.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.
J. Invest. Dermatol. 97, 980-984, 1991
A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polymerase chain reaction
A:Reference number: A54328; MUID:92084969; PMID:1748823
A:Accession: A54328
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-110 <PH2>
C:Genetics:
A:Gene: COL1A2
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:1145-1373/Domain: Fibrillar collagen carboxyl-terminal homology <FC>

QY	1	GSBEPEGVRCPPGPPAGAAGPADGCEPGAAGADGAPGAGPFPFGARPGSPGE	60
DB	280	GPAGPAGPREVGLUFGUSGVPVPCNTNLTKAKATGLGVAGAPGLGPRGIPGPA	339
QY	61	GPCGFPKGDSPGAPGSKDGTGAKCEPGFVGVEGPPGAPGEEKPGARGEPTGLP	120

Query Match 62.4%; Score 1915; DB 1; Length 1373;
Best Local Similarity 64.2%; Pred. No. 3.7e-86;
Matches 349; Conservative 45; Mismatches 150; Indels 0; Gaps 0;


```

Db      340 GAAGATGARGLVGEPGAGKSGESGNKGEPSVGAQGGPPGSGERGKSGSPGEASAGPA 399
QY      121 GPPGGERGGSGRFPAGDAGVAGFKGPAERGGSPGAPAGPKGSPEGAGRPGEAGLPAGKGLT 180
Db      400 GPPGLRGSQSRGLPCADGRAGVMGPPGNGSTGTGAGIRGPNAGDAGRPGEGLMGRGLP 459
QY      181 GSPGSGPPGKGTGPPGAGDGRDGPDPGPPGARGENGVGPPGPKAAGEPGKAGRGVVP 240
Db      460 GSPGNVPSGKEGFGVGLPIDGRPGTIGPAGPRGEAGNTGFPFGKPGSDGPKPGERGHP 519
QY      241 GPPGAVGPAGKGEAGAGAPPGGPAGPAGERGGEPAGSPGFGPLPGPAGPGEAGKPGEE 300
Db      520 GLAGARGAGPDNNNGAQPPFGPGVGGKGBGFPAGFPFQGLPGPSTTGEVKKPGR 579
QY      301 GVPDLAGAPSPGARGEPGPPGARGVEGPPGAPGAGDAGCGDCAKADGADGAPGAPSE 360
Db      580 GLPGEFGLPCAPGPRGERGTGPGSGAAGSPGTGSRGSPSCARGPDPNGKGEAGAVAPGSA 639
QY      361 GATGLSGMPGCEGAAGLPKPKDRGDAGPKGADGSGPKDGVRLGTGPIPGPPGAPGAPDK 420
Db      640 GASGPGGLPGERGAAGTGGKGBKGTGLRGDTGTGTRDGAAGTGAAGAPGAPGASGDR 699
QY      421 GESGSPGAPGPTGARGAGDRGPPGPPGAPGAGFPAGFPAGDGEFGAKGEPDAGAKGDAGPP 480
Db      700 GEAGAAGSPGAPGPRGSPGERGEVGPAGPAGNFGAPGAGAAQPGAKGKGTGPKGENGIV 759
QY      481 GPAGPAGPPGPIGDVAGPAGKAGGARGAGPPGATGFPGAGAGRVGPPGSPGADGPPGPPGA 540
Db      760 GPTGSGVGAAGSPGPPGPPGVRSGDGGPPGPTGFPAGAGRTGPPGPPGAGPPGPGAA 819
QY      541 KKEG 544
Db      820 KKEG 823

```

```

RESULT 14
I49607
collagen type V alpha 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49607
R:Andrikopoulos, K.; Suruki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A:Reference number: I49607; MUID:93214071; PMID:1297453
A:Accession: I49607
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1497 <RES>
A:Cross-references: GB:I02918; NID:9309180; PIDN:AAA37440.1; PID:9309181
C:Gene: Col5a-2
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <VWC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```

```

Query Match      62.2%; Score 1910; DB 2; Length 1497;
Best Local Similarity 63.8%; Pred. No. 6.9e-86;
Matches 347; Conservative 39; Mismatches 158; Indels 0; Gaps 0;

QY      1 GSEGEVGRGPPGPPGAGAGPAGDPCADGEPGAKGADGAPGAGPAGPAGPAGPAGPAGPAGP 60
Db      394 GPEGPGQGRGTGPPGAGSGQLPGAVGTDTGPKGATGAGTSGPGLAGPAGPAGPAGP 453
QY      61 GPGGPPGKGSBFGPAGSGKDTGAKGEPGVPVGGPPGAPGEGKPGARGPAGPAGTGLP 120
Db      454 GSTGPGQTGSGDGGPVGFGGEGAGKGPGEPHGIQGPFGPPGEGRGRGPPGTPGTP 513
QY      121 GPPGEGGPPGSRGPPGADGVAGPAGPAGERSGPPAGPKSGPAGPAGPAGPAGPAGPAGKGLT 180
Db      514 GPMERGAPGMRGPPGSDGLPGPKAQGERGPPVSSGPKGQGGPPGPPGPPGPPGAGLIT 573
QY      181 GSPGSGPPDGKTPGPPGAGEDGRFPFGPPGARGAGVMGFPFGKGAAGPFGKAGGERGVP 240

```

```

Db      574 GNPVQVQPEKGLFLGAPGEDGRGPPGSGIGIRGQPGSMGVPPGKSSGDLGRPGEAGNA 633
QY      241 GPPGAVGPAGKDEAGAGGPPGAPGAGERGEGPAGSPGFEGLPGPAGPAGPAGPAGPAGP 300
Db      634 GVPQRGAPGKDEGVGSPGVGPPGLAGERGEAGPPTGFGQLPGPPGPPGSGGKAGDQ 693
QY      301 GVPDCLGAPGSPGARGEPGPPGARGVEGPPGAPGAGDAGPDDGAKGADGAPGAPGSE 360
Db      694 GVPGEPAVGPGLGPRGRGNPGERGEGITGLPGEKMGAGHGDGPPGKGNPPTGTTGDT 753
QY      361 GAPLEGMPGEGGAAGLPKPKDRGDAGPKGADGSPGKDGVRGLTGTGPIGPPGAPGAPGDK 420
Db      754 GPPGLQGWPGERGTAGTTPGPKDRGGTGEKAGSTAGNDGARGLPGLGPPGAGLLGAP 813
QY      421 GESGSPGAPGPTGARGAGDRGPPGPPGAPGAGFPAGDGEFGAKGEPDAGAKGDAGP 480
Db      814 GEPGRLGVPGPPGSRGNPGRGENGPTGAVGFAGPQSGDQPGVKGFBGPPGQKGDAGSP 873
QY      481 GPAGPAGPPGPIGDVAGPAGKAGGARGSPGATGFPGAGAGRVGPPGSPGADGPPGPPGPA 540
Db      874 GPPGLASGPPGHPGHPGVLGKGRGTGTPGATGFPGAGRVGPPGPPGAPGAPGAPGAP 933
QY      541 KKEG 544
Db      934 KKEG 937

```

```

RESULT 15
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 13-Aug-1999
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOA>
A:Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322
R:Metzger, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; NID:950476; PIDN:CAA41048.1; PID:950477
C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 2
58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence #status predicted <SIG>
F:45-154/Domain: propetide #status predicted <PRO>
F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```

```

Query Match      62.1%; Score 1908; DB 2; Length 1464;
Best Local Similarity 63.3%; Pred. No. 8.5e-86;
Matches 346; Conservative 37; Mismatches 158; Indels 6; Gaps 2;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 7, 2004, 16:24:19 ; Search time 18 Seconds
(without alignments)
1573.676 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGFGEYGRGPPGPPGAGA.....PPSGDAGPPGPPGACKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1460	1	CA11_HUMAN
2	2940	95.8	1460	1	CA11_CANFA
3	2857	93.1	1453	1	CA11_MOUSE
4	2775	90.4	1453	1	CA11_CHICK
5	2702	71.7	1418	1	CA12_HUMAN
6	2192	71.4	1459	1	CA12_MOUSE
7	2151	70.1	671	1	CA11_RAT
8	2076.5	67.6	779	1	CA11_BOVIN
9	1949	63.5	1364	1	CA21_BOVIN
10	1934.5	63.0	1262	1	CA13_CHICK
11	1934	63.0	1372	1	CA21_RAT
12	1921.5	62.6	1496	1	CA25_HUMAN
13	1915	62.4	1372	1	CA21_MOUSE
14	1905.5	62.1	1049	1	CA13_BOVIN
15	1901	61.9	1466	1	CA13_HUMAN
16	1899	61.9	1464	1	CA13_MOUSE
17	1898	61.8	1366	1	CA21_CANFA
18	1880	61.2	1366	1	CA21_HUMAN
19	1834.5	59.8	1355	1	CA21_RANCA
20	1822	59.3	1362	1	CA21_CHICK
21	1789	58.3	1356	1	CA21_ONCMY
22	1496	48.7	747	1	CA12_BOVIN
23	1454.5	47.4	1838	1	CA15_HUMAN
24	1449	47.2	1806	1	CA15_HUMAN
25	1447	47.1	1745	1	CA35_HUMAN
26	1432	46.6	1804	1	CA15_MOUSE
27	1423.5	46.4	1736	1	CA2B_HUMAN
28	1414.5	46.1	1736	1	CA2B_MOUSE
29	1410	45.9	1027	1	CAFF_RIFPA
30	1387	45.2	1685	1	CA54_HUMAN
31	1375.5	44.8	2944	1	CA17_HUMAN
32	1372	44.7	1763	1	CA24_ASCSU
33	1357.5	44.2	1690	1	CA44_HUMAN

RESULT 1

ID	CALL_HUMAN	STANDARD	PRT	1464 AA
AC	P02452; Q14037; Q15176;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Collagen alpha 1(I) chain precursor.			
GN	COL1A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE OF 1-472 FROM N.A.			
RX	MEDLINE=89025644; PubMed=3178743;			
RA	Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,			
RA	Jaenisch R., Prockup D.J.;			
RT	"Structure of a full-length cDNA clone for the prepro alpha 1(I)			
RT	chain of human type I procollagen.";			
RL	Biochem. J. 253:919-922(1988).			
RN	[2]			
RP	SEQUENCE OF 1-181 FROM N.A.			
RX	MEDLINE=84270697; PubMed=6462220;			
RA	Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,			
RA	Myers J., Williams C., Ramirez F.;			
RT	"Human pro alpha 1(I) collagen gene structure reveals evolutionary			
RT	conservation of a pattern of introns and exons.";			
RL	Nature 310:337-340(1984).			
RN	[3]			
RP	SEQUENCE OF 162-301.			
RX	TISSUE=Skin;			
RC	MEDLINE=71038625; PubMed=5529814;			
RX	Click E.M., Bornstein P.;			
RA	"Isolation and characterization of the cyanogen bromide peptides from			
RT	the alpha 1 and alpha 2 chains of human skin collagen.";			
RL	Biochemistry 9:4699-4706(1970).			
RN	[4]			
RP	SEQUENCE OF 263-268.			
RX	TISSUE=Skin;			
RC	MEDLINE=71001508; PubMed=4319110;			
RX	Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;			
RA	"A comparative study of glycopeptides derived from selected			
RT	vertebrate collagens. A possible role of the carbohydrate in fibril			
RT	formation.";			
RL	J. Biol. Chem. 245:5042-5048(1970).			
RN	[5]			
RP	SEQUENCE OF 425-1464 FROM N.A.			
RX	MEDLINE=84080385; PubMed=6689127;			
RA	Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,			
RA	Prockup D.J.;			
RT	"Nucleotide sequences of complementary deoxyribonucleic acids for the			
RT	pro alpha 1 chain of human type I procollagen. Statistical evaluation			
RT	of structures that are conserved during evolution.";			
RL	Biochemistry 22:5213-5223(1983).			
RN	[6]			

Q28083 bos taurus
Q01955 homo sapien
Q07092 homo sapien
P02462 homo sapien
P17140 caenorhabdi
P02463 mus musculus
P08572 homo sapien
P39060 homo sapien
Q28247 canis famil
Q14050 homo sapien
Q14993 homo sapien
Q05722 mus musculus

34 1356 44.2 911 1 CALB_BOVIN
35 1344.5 43.8 1670 1 CA34_HUMAN
36 1339.5 43.6 1803 1 CA1F_HUMAN
37 1320 43.0 1869 1 CA14_HUMAN
38 1310.5 42.7 1758 1 CA24_CABEL
39 1310 42.7 1669 1 CA14_MOUSE
40 1286.5 41.9 1712 1 CA34_HUMAN
41 1281.5 41.7 1516 1 CA1H_HUMAN
42 1279.5 41.7 754 1 CA54_CANFA
43 1277 41.6 684 1 CA39_HUMAN
44 1276.5 41.6 1143 1 CA1I_HUMAN
45 1275 41.6 921 1 CA19_MOUSE

ALIGNMENTS

RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelin R.E.;
RT "Regulatory elements in the first intron contribute to
transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
associated collagen (type IX), and network-forming collagen (type X)
cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalgleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelin R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II ARG-569.
RX MEDLINE=8722295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
arginine for glycine at residue 391 of the alpha 1(I) chain of type I
collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=9009313; PubMed=2794057;
RA Starnan B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
cysteine in the triple helical domain of the pro alpha 1(I) chains of
type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1206-1214(1989).
RN [25]
RP VARIANT OI-II CYS-422.
RX MEDLINE=90062058; PubMed=2511192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
variant of osteogenesis imperfecta minimally destabilizes the triple
helix of type I procollagen. The effects of glycine substitutions on
thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19695(1989).
RN [23]
RP VARIANT OI-II CYS-1082.
RX MEDLINE=89109573; PubMed=2913053;
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT "A lethal variant of osteogenesis imperfecta has a single base
mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
chain of type I procollagen. The asymptomatic mother has an
unidentified mutation producing an overmodified and unstable type I
procollagen.";
RL J. Clin. Invest. 83:574-584(1989).
RN [24]
RP VARIANTS OI CYS-272; CYS-704 AND CYS-896.
RX MEDLINE=9009313; PubMed=2794057;
RA Starnan B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
cysteine in the triple helical domain of the pro alpha 1(I) chains of
type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1206-1214(1989).
RN [25]
RP VARIANT OI-II CYS-422.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
of the alpha 1 chain to cysteine and destabilizes the triple helix in
a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
RN [17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lemande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
chain in lethal perinatal osteogenesis imperfecta. Demonstration of
the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
RN [18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218626; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
1(I) chain of type I collagen in a patient with mild dominantly
inherited osteogenesis imperfecta.";
RL Mol. Biol. Med. 5:197-207(1988).
RN [19]
RP VARIANT OI-II VAL-434.
RX MEDLINE=89255493; PubMed=2470760;
RA Patterson E., Smiley E., Bonadio J.;
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
mutation.";
RL J. Biol. Chem. 264:10083-10087(1989).
RN [20]
RP VARIANT OI-IV SER-1010.
RX MEDLINE=89308591; PubMed=2745420;
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL J. Biol. Chem. 264:11893-11900(1989).
RN [21]
RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX MEDLINE=89380165; PubMed=2777764;
RA Lemande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
RN [22]
RP VARIANT OI SER-1022.
RX MEDLINE=90062058; PubMed=2511192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
variant of osteogenesis imperfecta minimally destabilizes the triple
helix of type I procollagen. The effects of glycine substitutions on
thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19695(1989).
RN [23]

Query Match 96.6%; Score 2966; DB 1; Length 1464;
Best Local Similarity 95.6%; Pred. No. 2.5e-113;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGGVGRGPPGPPGAGAGPAGDPCADGEPGAKGADGAPGTAGAPGPPGARGPSGPE 60
DB 362 GSEGGVGRGPPGPPGAGAGPAGNADQOPGAKGANGAPGTAGAPGPPGARGPSGPPQ 421
QY 61 GFGGPPGKGSGEFGAPGSGKDTGAKGEPGVGVGPPGPPGAGGEGKPGARGGEPGPTGLP 120
DB 422 GFGGPPGKGSGEFGAPGSGKDTGAKGEPGVGVGPPGPPGAGGEGKPGARGGEPGPTGLP 481
QY 121 GPPGREGGSGRFGGAGGVAGPKGAPAGERSGPPGAGPKGSPGARGRGEAGLPGAKGLT 180
DB 482 GPPGREGGSGRFGGAGGVAGPKGAPAGERSGPPGAGPKGSPGARGRGEAGLPGAKGLT 541
QY 181 GSPGSGPDGKTGTPGPPGAGEDGPPGPPGPPGARGGAGVWGPFGPKGAAGEGPKAGRGVP 240
DB 542 GSPGSGPDGKTGTPGPPGAGEDGPPGPPGPPGARGGAGVWGPFGPKGAAGEGPKAGRGVP 601
QY 241 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGGGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 602 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGGGPPGPPGPPGPPGPPGPPGPPGPPGPPG 661
QY 301 GVPGLGAPGSGGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 662 GVPGLGAPGSGGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 721
QY 361 GAGLEGVMPGREGAAGLPGPKDRDAGPKGADGSGPKDGVGRGLTGTPTGPPGPPGAGPGDK 420
DB 722 GAGLEGVMPGREGAAGLPGPKDRDAGPKGADGSGPKDGVGRGLTGTPTGPPGPPGAGPGDK 781
QY 421 GESGSGGAGPTGARGAGDGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB 782 GESGSGGAGPTGARGAGDGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 841
QY 481 GPAGPAGPPGPTGVDGAPGAKGARGGAGPPGATGPPGGAAGVGGPPGPPGPPGPPGPPG 540
DB 842 GPAGPAGPPGPTGVDGAPGAKGARGGAGPPGATGPPGGAAGVGGPPGPPGPPGPPGPPG 901
QY 541 GKEG 544
DB 902 GKEG 905

RESULT 2
CAL1 CANFA
ID CAL1 CANFA STANDARD; PRT; 1460 AA.

AC Q9X5U7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OC NCBI_TaxID=9615;
RN [1]_TAXID=9615;
RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
RC TISSUE=Skin;
RX MEDLINE=21023337; PubMed=11147834;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a
heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case
of canine osteogenesis imperfecta."
RL Arch. Biochem. Biophys. 384:37-46(2000).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -!- SUBUNIT: trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta (OI).
CC -!- SIMILARITY: Contains 1 VWFC domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF153062; AAD34619.1; -.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR008985; Fib_collagen_C.
DR InterPro: IPR002181; Fibrinogen_C.
DR InterPro: IPR001007; VWFC_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR ProDom: PD000007; Clg_helix; 2.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWFC; 1.
DR PROSITE: PS01208; VWFC 1; 1.
DR PROSITE: PS0184; VWFC 2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Disease mutation.
FT SIGNAL 1 22 BY SIMILARITY.
FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 34 92
FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.
FT DOMAIN 1189 1214 NONHELICAL REGION (C-TERMINAL).
FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 1361 1361 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 208 208 G -> A (in OI; severe).
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 95.8%; Score 2940; DB 1; Length 1460;
Best Local Similarity 94.3%; Pred. No. 2.8e-112;
Matches 513; Conservative 24; Mismatches 7; Indels 0; Gaps 0;

QY 1 GSEGGVGRGPPGPPGAGAGPAGDPCADGEPGAKGADGAPGTAGAPGPPGARGPSGPE 60
DB 358 GSEGGVGRGPPGPPGAGAGPAGNADQOPGAKGANGAPGTAGAPGPPGARGPSGPPQ 417
QY 61 GPGGPPGKGSDEGAPGAPGSKDTGAKGEPGVGVGPPGPPGAGGEGKPGARGGEPGPTGLP 120
DB 418 GPGGPPGKGSDEGAPGAPGSKDTGAKGEPGVGVGPPGPPGAGGEGKPGARGGEPGPTGLP 477
QY 121 GPPGREGGSGRFGGAGGVAGPKGAPAGERSGPPGAGPKGSPGARGRGEAGLPGAKGLT 180
DB 478 GPPGREGGSGRFGGAGGVAGPKGAPAGERSGPPGAGPKGSPGARGRGEAGLPGAKGLT 537
QY 181 GSPGSGPDGKTGTPGPPGAGEDGPPGPPGPPGARGGAGVWGPFGPKGAAGEGPKAGRGVP 240
DB 538 GSPGSGPDGKTGTPGPPGAGEDGPPGPPGPPGARGGAGVWGPFGPKGAAGEGPKAGRGVP 597
QY 241 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGGGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 598 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGGGPPGPPGPPGPPGPPGPPGPPGPPGPPG 657
QY 301 GVPGLGAPGSGGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 658 GVPGLGAPGSGGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 717
QY 361 GAGLEGVMPGREGAAGLPGPKDRDAGPKGADGSGPKDGVGRGLTGTPTGPPGPPGAGPGDK 420
DB 718 GAGLEGVMPGREGAAGLPGPKDRDAGPKGADGSGPKDGVGRGLTGTPTGPPGPPGAGPGDK 777

[illegible]

```

RESULT 3
CALL_MOUSE STANDARD; PRT; 1453 AA.
ID CALL_MOUSE STANDARD; PRT; 1453 AA.
AC F11087; Q60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COL1A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
[1]
RN SEQUENCE FROM N.A.
RP
RC STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Killian J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen.";
RL Matrix Biol. 14:593-595(1995).
[2]
RN SEQUENCE OF 518-1128 FROM N.A.
RP
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein.";
RL Gene 39:311-312(1985).
[3]
RN SEQUENCE OF 735-1130 FROM N.A.
RP
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371(1982).
[4]
RN SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RP
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69(1981).
[5]
RN SEQUENCE OF 1442-1453 FROM N.A.
RP
RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773(1988).
CC -1- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- SIMILARITY: Contains 1 VWFC domain.

```

RA	Moison J.M., McCarthy B.J.;	
RT	"Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:	
RT	"evidence for insertions or deletions in gene coding sequences.";	
RT	DNA 1:59-69(1981).	
RN	[5]	
RP	SEQUENCE OF 1442-1453 FROM N.A.	
RA	MEDLINE=88124276; PubMed=3340560;	
RX	Mooslehner K., Harbers K.;	
RT	"Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size	
RT	of the 3'-untranslated region.";	
RL	Nucleic Acids Res. 16:773-773(1988).	
CC	-1- FUNCTION: Type I collagen is a member of group I collagen	
CC	(-1- (fibrillar forming collagen).	
CC	-1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.	
CC	-1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and	
CC	bones. In bones the fibrils are mineralized with calcium	
CC	hydroxyapatite.	
CC	-1- PTM: prolines at the third position of the tripeptide repeating	
CC	unit (G-X-Y) are hydroxylated in some or all of the chains.	
CC	-1- SIMILARITY: Contains 1 VWFC domain.	
CC	-----	

[illegible]

RESULT 4

CALL-CHICK	STANDARD;	PRT; 1453 AA.
ID		
AC	P02457;	
DT	21-JUL-1986 (Rel. 01, Created)	
DT	01-OCT-1989 (Rel. 12, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Collagen alpha 1(I) chain precursor.	
GN	COL1A1.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;	
OC	Gallus.	
OX	NCBI_TaxID=9031;	
[1]	SEQUENCE OF 1-153 FROM N.A.	
RP	MEDLINE=88056316; PubMed=3678834;	
RX	Finer M.H.; Boedtker H.; Doty P.;	
RA	"Construction and characterization of cDNA clones encoding the 5' end	
RT	of the chicken pro alpha 1(I) collagen mRNA.";	
RL	Gene 56:71-78(1987).	
[2]	SEQUENCE OF 1-144 FROM N.A.	
RP	MEDLINE=88007542; PubMed=2820966;	
RX	Finer M.H.; Abo S.; Gerstenfeld L.C.; Boedtker H.; Doty P.;	
RA	"Unusual DNA sequences located within the promoter region and the	
RT	first intron of the chicken pro-alpha 1(I) collagen gene.";	
RL	J. Biol. Chem. 262:13323-13332(1987).	
[3]	SEQUENCE OF 152-1187.	
RP	MEDLINE=82231995; PubMed=7093229;	
RX	Highberger J.H.; Corbett C.; Dixit S.N.; Yu W.; Seyer J.M.;	
RA	Kang A.H.; Gross J.;	
RA	"Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the	
RT	complete primary structure of the helical portion of the chick skin	
RL	collagen alpha 1(I) chain.";	
RL	Biochemistry 21:2048-2055(1982).	
[4]	SEQUENCE OF 1200-1205.	
RP	MEDLINE=72243016; PubMed=5047697;	
RX	Eyre D.R.; Glimcher M.J.;	
RA	"Evidence for a previously undetected sequence at the carboxyterminus	
RT	of the alpha 1 chain of chicken bone collagen.";	
RL	Biochem. Biophys. Res. Commun. 48:720-726(1972).	
[5]	SEQUENCE OF 981-1453 FROM N.A.	
RP	MEDLINE=81160715; PubMed=6927845;	
RX	Fuller F.; Boedtker H.;	
RA		

Sequence determination and analysis of the 3' region of chicken pro-alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids including the carboxy-terminal propeptide sequences.;

Biochemistry 20:996-1006(1981).

[6]

SEQUENCE OF 1311-1453 FROM N.A.

MEDLINE-80134546; PubMed-6987088;

Shewalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T., Pastan I., Decrombrughe B., Flaetzek P.P., Olsen B.R.;

"Nucleotide sequence of a collagen cDNA-fragment coding for the carboxyl end of pro alpha 1(I)-chains.;"

FEBS Lett. 111:61-65(1980).

PFES

-I- FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).

-I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.

-I- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and bones. In bones the fibrils are mineralized with calcium hydroxyapatite.

-I- PTM: Proline residues at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains. Pro-1153 is the only 3-hydroxypro and the only hydroxylated proline in position X.

-I- SIMILARITY: Contains 1 VWFC domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no commercial modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

EMBL; M17839; AAA48704.1; -.
EMBL; M17838; AAA48704.1; JOINED.
EMBL; V00401; CAA23695.1; -.
EMBL; M10571; AAA48671.1; ALT_SEQ.
EMBL; M17607; AAA48672.1; -.
PIR; A27179; A27179.
PIR; I50629; I50629.
InterPro; IPRO08161; Clg_helix.
InterPro; IPRO08160; Collagen.
InterPro; IPRO00885; Fib_collagen_C.
InterPro; IPRO02181; Fibrinogen_C.
InterPro; IPRO01007; VWF_C.
Pfam; PF01410; COLFI; 1.
Pfam; PF01391; Collagen; 18.
Pfam; PF00093; VWC; 1.
ProDom; PD000007; Clg_helix; 2.
ProDom; PD002078; Fib_collagen_C; 1.
SMART; SMO00338; COLFI; 1.
SMART; SMO0214; VWC; 1.
PROSITE; PS01208; VWF_C; 1.
PROSITE; PS0184; VWF_C; 2; 1.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
SIGNAL 1 22
PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
DOMAIN 31 89 VWC.
MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
MOD_RES 1153 1153 HYDROXYLATION.
CONFLICT 1187 1187 F -> L (IN REF. 5).
CONFLICT 1441 1441 Q -> H (IN REF. 6).
SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
90.4%; Score 2775; DB 1; Length 1453;
ary Match 89.0%; Pred. No. 1.2e-105; Indels 0; Gaps 0;
ches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

Query Match 90.4%; Score 2775; DB 1; Length 1453;
Best Local Similarity 89.0%; Pred. No. 1.2e-105;
Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
RN [14]
RP VARIANT OSTEOARTHRTIS CVS-650.
RX MEDLINE=90370826; PubMed=1975693;
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia".
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN [15]
RP VARIANT OI-IV VAL-717.
RX MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [16]
RP VARIANT OSTEOARTHRTIS CVS-650.
RX MEDLINE=91086471; PubMed=1985108;
RA Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
RN [17]
RP VARIANT HYPOCHONDROGENESIS GLU-984.
RX MEDLINE=93054548; PubMed=1429602;
RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RA Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly953-->Glu) in the collagen alpha
RT 1(I) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
RN [18]
RP VARIANT HYPOCHONDROGENESIS SER-705.
RX MEDLINE=92262484; PubMed=1374506;
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA Ramirez F., Vitale E., Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN [19]
RP VARIANT WS-II ASP-198.
RX MEDLINE=93304428; PubMed=9317498;
RA Koerikko J., Ritvanemi P., Haataja L., Kaaerlaeinen H.,
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy)".
RL Am. J. Hum. Genet. 53:55-61(1993).
RN [20]
RP VARIANT SEMD CYS-840.
RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
RA Eyre D.R.;
RT "A dominant mutation in the type II collagen gene (COL2A1) produces
RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
RL Am. J. Hum. Genet. 53:A209-A209(1993).
RN [21]
RP VARIANT OSTEOARTHRTIS CVS-650.
RX MEDLINE=93282819; PubMed=8507190;
RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;
RT "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
RN [22]
RP VARIANT SEMD ARG-285.
RX MEDLINE=93252400; PubMed=8486375;
RA Vakkula M., Ritvanemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
RA Peltonen L.;
RT "A mutation in the amino-terminal end of the triple helix of type II
RT collagen causing severe osteochondrodysplasia.";
RL Genomics 16:282-285(1993).

[23]
RN VARIANT SEDC CYS-206.
RX MEDLINE=94063862; PubMed=8244341;
RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT family with an Arg75-->Cys mutation in the procollagen type II gene
RT (COL2A1)".
RL Hum. Genet. 92:499-505(1993).
RN [24]
RP VARIANT SEDC CYS-920.
RX MEDLINE=93315508; PubMed=8325895;
RA Chan D., Taylor T.K.F., Cole W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. Biol. Chem. 268:15238-15245(1993).
RN [25]
RP VARIANT SEDC SER-1128.
RX MEDLINE=93140139; PubMed=8423604;
RA Cole W.G., Hall R.K., Rogers J.G.;
RT "The clinical features of spondyloepiphyseal dysplasia congenita
RT resulting from the substitution of glycine 997 by serine in the alpha
RT 1(II) chain of type II collagen.";
RL J. Med. Genet. 30:27-35(1993).
Query Match 71.7%; Score 2202; DB 1; Length 1418;
Best Local Similarity 71.9%; Pred. No. 1.3e-82;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
QY 1 GSEGGVYGERGPPGAGAGPAGDPCADGSPGAKGADGAPCIAGAPFPGARPSGPE 60
DB 315 GPEGAGGPRGEGTTPGSPGAGASGNTGDTGIPGAKGAGAPCIAGAPFPGARPSGPE 374
QY 61 GPGGPPGPKGDSGEPGAPGSKDGTGAKGEPGVPVGGPPGAPGKGPARGEPGPTGUP 120
DB 375 GATGTLGPKGTGKPKIAGFKGEGQPKGEPGAPGQAGPAGPAGERGKRGARGEPGVGPI 434
QY 121 GPRGRRGSGRGFCAGDAGVAGKGPAGRGSPGAPGKSGPGEAGLPGAKGUT 180
DB 435 GPPGARGAPGNRGFPQDGLAGPKGAPGERSGLAGPKANGDPGRPGEPGLPGAGUT 494
QY 181 GSPGSPGPDGKTGPPGAGEDGRPPGPPGARGAGVGMFFGPKGAAGEPKKAGERGVP 240
DB 495 GRFGDAGPQGVGSPGAPGEDGRPPGPPGQARGQPGVGMFFGPKGANGEPKAGEKGLP 554
QY 241 GPPGAVGPKDGEAGAGGPPGAPGAPGREGGEPGAPSGFGLPGPAGPGEAGKPGEE 300
DB 555 GAPGLGKGFCDGEGTGAEGPPGAPGAPGREGGQAPSGFGQLPGPPGPPGEGKPGDQ 614
QY 301 GVPGLGAPGPGCARGEPGPPGREGVGGPPGAPGPPGADGAPGDDGAKGDAGAPGSE 360
DB 615 GVPGEAGAPGLVGRGERGPPGERSPGAQGLQGPRGLFTGTGDKGASGAPGPGAQ 674
QY 361 GAPLEGMPGERGAAGLPKPKDRGDAGPKGADGSPKDGVRGLTGTGTPGPPGAPGDK 420
DB 675 GPPGLQGMFGERGAAGIAGFKGDRGVGKGPAGPKDGGRLTGTGTPGPPGAPGSEK 734
QY 421 GESGSPGAPGTPGARGAPGDRGSPGPPGAPGAPGPPGADGEPGAKGEPDAGAKGDAGPP 480
DB 735 GEVGPFPAGSAGAPGAPGERTGTPGTSGLAGPPGADGQPGAKGEQGEAKGDAGAP 794
QY 481 GPAGPAGPPGPIGDVGPAGKAGSAGPPGATGFPGAAGRVPSPGSDAGPPGPPGPA 540
DB 795 GPQGPSGAPGPPGTPGTGTPGKARGAQGPPGATGFPGAAGRVPSPGSDAGPPGPPG 854
QY 541 GKEG 544
DB 855 GKDG 858
RESULT 6
CA12_MOUSE

hydroxyapatite.
-!- PTM: Proline residues at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to the oxygen atom of a post-translationally added hydroxyl group.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@sib-sib.ch.

EMBL; M1432; AAA40832.1; AUT SEQ.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF C.
DR ProDom; PD000007; Clg_helix; 1.
DR PROSITE; PS01208; WVF C 1; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT MOD_RES 9 9 CONVERTED TO AN ALDEHYDE GROUP THAT IS INVOLVED IN CROSS-LINKING.
FT FT 28 28 HYDROXYLATION (PROBABLE).
FT MOD_RES 31 31 HYDROXYLATION (PROBABLE).
FT MOD_RES 34 34 HYDROXYLATION (PROBABLE).
FT MOD_RES 43 43 HYDROXYLATION (PROBABLE).
FT MOD_RES 46 46 HYDROXYLATION (PROBABLE).
FT MOD_RES 49 49 HYDROXYLATION (PROBABLE).
FT MOD_RES 103 103 HYDROXYLATION.
FT MOD_RES 103 103 O-LINKED (GAL...).
FT CARBOHYD 424 424 HYDROXYLATION (PROBABLE).
FT MOD_RES 547 547 HYDROXYLATION (PROBABLE).
FT FT 567 568
FT NON_CONS 641
FT DOMAIN
FT FT
FT FT
SQ SEQUENCE 671 AA; 60615 MW; 9DC3114204AC4918 CRC64;

Query Match 70.1%; Score 2151; DB 1; Length 671;
Best Local Similarity 77.5%; Pred. No. 9.1e-81;
Matches 382; Conservative 27; Mismatches 36; Indels 48; Gaps 2;

QY 1 GSEGGCVGRGEPGPPGAGAGPAGPGADGEPGAKGADGAPGIAGAPGPGARGPSGPE 60
DB 200 GSEGGCVGRGEPGPPGAGAGPAGPGADGQFGKANGANGAPGIAGAPGPGARGPSGP 259

QY 61 GPGGPPGPKDSEGPAGPSGKGTGAKGEPGVGVEGPPGAPGEEGKGARGEGPTGLP 120
DB 260 GPSGAPGPKNSGEPGAPGNKGTGAKGEPGAVGQPPGAPGEEGKRGARGEPGSGLP 319

QY 121 GPPEERGSGSRFPFGADGVAGPKGPAGERGSGPPGAPKSGPGEAGLPGAKGLT 180
DB 320 GPPEERGSGSRFPFGADGVAGPKGPAGERGSGPPGAPKSGPGEAGLPGAKGLT 379

QY 181 GSPGSPGPDGKTGPPGPGAGEDGPPGPPGARGEGAGVWGFPQPKGAAGPFGKAGERGVP 240
DB 380 GSPGSPGPDGKTGPPGPGAGZGPPGPPGARGEGAGVWGFPQPKGTAGEPFGKAGERGVP 439

QY 241 GPGAVGPKGDKGAEAGGPPGPPGAPAGERGEEGPGSPGEGULPGAGPPGAGKGPGE 300
DB 440 GPGAVGPKGDKGAEAGGPPGPPGAPAGERGEEGPGSPGEGULPGAGPPGAGKGPGE 499

QY 301 GVPGLGAPGFGARGSGPFGPFGERGVEGPPGAPGPGADGAPGDDGAKGDAGAPGPGSE 360
DB 500 GVPGLGAPGFGARGSGPFGERGVEGPPGAPGPGADGAPGDDGAKGDAGAPGPGSQ 559

QY 361 GAGLEGNPGERGAAGLPGPKGDRGDAGPKGADGSPGKDVGRGLTGPIGPPGAPGAGDK 420
DB 560 GAGLQGMGLZ-----GPPGPPGSGSGZ 583

RESULT 9
CA21 BOVIN STANDARD; PRT; 1364 AA.
ID C02455; 062649;
DT 21-JUL-1986 (Rel. 01, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1] SEQUENCE FROM N.A.
RP TISSUE=Aorta;
RC MEDLINE=98290219; PubMed=9628255;
RA Shirai T., Hattori S., Sakaguchi M., Inouye S., Kimura A., Ebihara T.,
RA Irie S., Nagai Y., Hori H.;
RT "The complete cDNA coding sequence for the bovine proalpha2(I) chain
RT of type I procollagen.";
RL Matrix Biol. 17:85-88(1998).
RN [2]
RP SEQUENCE OF 80-98.
RC TISSUE=Skin;
RX MEDLINE=75036115; PubMed=4609475;
RA Fietzek P.P., Breitkreutz D., Kuehn K.;
RT "Amino acid sequence of the amino-terminal region of calf skin
RT collagen.";
RL Biochim. Biophys. Acta 365:305-310(1974).
RN [3]
RP SEQUENCE OF 95-415, AND REVISION.
RC TISSUE=Skin;
RX MEDLINE=76091874; PubMed=173531;
RA Fietzek P.P., Rexrodt F.W.;
RT "The covalent structure of collagen. The amino-acid sequence of
RT alpha2-CB4 from calf-skin collagen.";
RL Eur. J. Biochem. 59:113-118(1975).
RN [4]
RP SEQUENCE OF 416-445.
RC TISSUE=Skin;
RX MEDLINE=75008198; PubMed=4412529;
RA Fietzek P.P., Furthmayr H., Kuehn K.;
RT "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
RT and pig-skin collagen.";
RL Eur. J. Biochem. 47:257-261(1974).
RN [5]
RP SEQUENCE OF 446-481.
RC TISSUE=Skin;
RX MEDLINE=75059250; PubMed=4435743;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RT from calf skin collagen.";
RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
CC -! FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -! SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -! TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -! PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL; AB008683; BAA25171.1; ..
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 24
FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 80 1100 COLLAGEN ALPHA 2(I) CHAIN.
FT PROPEP 1101 1364 CARBOXYL-TERMINAL PROPEPTIDE.
FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 84 84 INVERTED TO AN ALDEHYDE GROUP THAT IS
FT MOD_RES 175 175 INVOLVED IN CROSS-LINKING (PROBABLE).
FT MOD_RES 196 196 HYDROXYLATION (PROBABLE).
FT MOD_RES 262 262 HYDROXYLATION (PROBABLE).
FT MOD_RES 307 307 HYDROXYLATION (PROBABLE).
FT MOD_RES 352 352 HYDROXYLATION (PROBABLE).
FT CONFLICT 157 157 V -> P (IN REF. 3).
FT CONFLICT 187 187 K -> T (IN REF. 3).
FT CONFLICT 211 211 T -> K (IN REF. 3).
FT CONFLICT 298 300 PGA -> AGP (IN REF. 3).
FT CONFLICT 423 424 AT -> TA (IN REF. 4).
SQ SEQUENCE 1364 AA; 129064 MW; 5593F4D6B9ED119A CRC64;
Query Match 63.5%; Score 1949; DB 1; Length 1364;
Best Local Similarity 65.3%; Pred. No. 1.9e-72;
Matches 355; Conservative 44; Mismatches 145; Indels 0; Gaps 0;
QY 1 GSEGEVGRGEPGPGPAGAGPAGGADGEGAGKAGDAGAPGPPGAGRGSGPE 60
DB 272 GPAGPAGRGVEGLGSLGPGVPPGANGLPAGKAGLPGVAGAPGLGPRGIPGV 331
QY 61 GPQGPFGKDSGEPGAPGSKDGTAKGEPGVGVGPPGPPGAGGEGKPGARGFPGTGLP 120
DB 332 GAAGATGARGLVGEPGPGAGSKGSGKNGEPGAVGPGPPGSGEGKRGSTGTFGAPGP 391
QY 121 GPPGRRGPGSRGPPGADGVAGPKGAGRGSGPAGPKGSPGEGAGPPGAGLPGAKGLT 180
DB 392 GPPGLRGNPGSRGLPGADGRAGVNGPAGSRGATGAPGVRFNGDSGRPPGLMGRGFP 451
QY 181 GSPGSPGDKTGTGPPGAGDGRGPPGPPGPPGAGGAGVGMGPPGKGAAGPPGKAGRGVP 240
DB 452 GSPGNIGPAGKEGPGVGLGIDGRPGIPGAPAGGEGPNIIGFPKGPSPGPKAGKGHA 511
QY 241 GPPGAVGPKDGEAGAGPPGPPGAGGERGEGPAGSPGEGPGLPGPAGPPGAGKGE 300
DB 512 GLAGARGAPGPDGNNGAQGGPGLQGQGGQGGPAGPPGFGPLPGPAGTAGGAKPGER 571
QY 301 GVRGLGAPGSPGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
DB 572 GIPGEGPLPGPAGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 631
QY 361 GAGLEGMPGEGRGAAGLPKPGKRGDAGPKGADGSPKGVRLGTGIPGPPGAGAPDK 420
DB 632 GPSGPGSLPGERGAAGIPGKGEGTGLRGDGGPGRDARGAPGAPGAIGAPGAGANDR 691
QY 421 GESGSPGAPGTGARGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB 692 GEAGPAGPAGPAGPRGSGRGERGVEGVPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 751
QY 481 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 540
DB 752 GPTGVGAAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 811
QY 541 KKEG 544
DB 812 KKEG 815

RESULT 10

CAL3 CHICK

ID -CAL3 CHICK STANDARD; PRT: 1262 AA.

AC B12105; P79758; P79759; Q90794; Q92029;

DT 01-OCT-1989 (Rel. 12, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE Collagen alpha 1(III) chain precursor (fragments).

GN COL3A1.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

ON NCBI_TaxID=9031;

RX [1]

RP SEQUENCE OF 1-886 FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=94266842; PubMed=8206952;

RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen."

RL J. Biol. Chem. 269:16443-16448(1994).

[2]

RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.

RX MEDLINE=84270656; PubMed=6547770;

RT Yamada Y., Liu G., Mudryj M., Obici S., de Crombrughe B.;

RA "Conservation of the sizes for one but not another class of exons in two chick collagen genes."

RL Nature 310:333-337(1994).

[3]

RP SEQUENCE OF 977-1262 FROM N.A.

RX MEDLINE=83220816; PubMed=6856474;

RT Yamada Y., Kuhn K., de Crombrughe B.;

RA "A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes."

RL Nucleic Acids Res. 11:2733-2744(1983).

-I- FUNCTION: Collagen type III occurs in most soft connective tissues along with type I collagen.

-I- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are linked to each other by interchain disulfide bonds. Trimers are also cross-linked via hydroxylysines.

-I- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

-I- SIMILARITY: Contains 1 VWFC domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U07973; AAA83407.1; --

DR EMBL; X00822; CAB52686.1; --

DR EMBL; X00823; CAB52686.1; JOINED.

DR EMBL; X00826; CA223397.1; ALT SEQ.

DR EMBL; X00825; CA223397.1; JOINED.

DR EMBL; X00827; CA223398.1; --

DR EMBL; X00828; CA223399.1; --

DR EMBL; X00830; CA225401.1; --

DR EMBL; X00831; CA225402.1; --

DR EMBL; K02302; RAD15299.1; --

DR EMBL; K02301; RAD15298.1; --

DR EMBL; M36662; RAA18519.1; ALT_SEQ.

PIR; A05269; A05269.

PIR; I50694; I50694.

DR InterPro; IPR008161; Clg_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR008085; Fib_collagen_C.

DR InterPro; IPR002181; Fibrinogen_C.

DR InterPro; IPR001007; VWFC_C.

DR ProDom; PD000007; Clg_helix_1.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM000038; COLFI; 1.

DR SMART; SM00214; VWC; 1.

DR PROSITE; PS01208; VWFC_1; 1.

DR PROSITE; PS0184; VWFC_2; 1.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Collagen protein; Collagen; Signal.

XN SIGNAL 1 23 POTENTIAL.

FT PROPEP 24 144 AMINO-TERMINAL PROPEPTIDE (BY SIMILARITY).

FT CHAIN 145 1003 COLLAGEN ALPHA 1(III) CHAIN.

FT PROPEP 1004 1262 CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).

FT DOMAIN 29 88 VWFC.

FT DOMAIN 145 164 NONHELICAL REGION (N-TERMINAL) (BY SIMILARITY).

FT DOMAIN 165 994 TRIPLE-HELICAL REGION (BY SIMILARITY).

FT DOMAIN 995 1003 NONHELICAL REGION (C-TERMINAL) (BY SIMILARITY).

FT NON_CONS 886 887 INTERCHAIN (BY SIMILARITY).

FT NON_CONS 922 923 INTERCHAIN (BY SIMILARITY).

FT DISULFID 994 994 HYDROXYLATION (BY SIMILARITY).

FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).

FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).

FT MOD_RES 859 859 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 1163 1163 E -> K (IN REF. 2).

FT CONFLICT 96 96 F -> S (IN REF. 3).

SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;

Query Match 63.0%; Score 1934.5; DB 1; Length 1262;

Best Local Similarity 63.8%; Pred. No. 7e-72; Indels 9; Gaps 1;

Matches 353; Conservative 31; Mismatches 160;

YQ 1 GSSEGEVGRGEPGPAGAGADPGDGADGEPKAKGADGAPGAFPPCAFGPSGP 60

DB 359 GASGNPGERGEFGPQGAGPQPQGGPGRAGSPGKGGMGSGIPGGPGGRLPGPP 418

YQ 61 GPGGPPPKGDGFPCAGSKGDKTCAKEFPVGVEGPPGAGEGKPGAGEGCPGLP 120

DB 419 GTSGNPFAKGTPEEPKNGAKDPCKPGERGENGTFFGARGPFGEGKRGANGEPQNGVP 478

YQ 121 GPPGERGGPGSRGFFCADGVAGKGPAGERSPGPAGPKGSFGAGRPGEAGLPCKAKLT 180

DB 479 GFPERGSGFPLGSLPGSNGLPCEKXPAGERSGPSPGSGPADRGDGGFLPMRGLP 538

YQ 181 GSPGSPGPKTKTPGPPAGEDRGPPPPGARGAGVMGPPGKAAGEGFKAGERGVP 240

DB 539 GTPGSPGSGKPFQPNQEPGRSGPPGPPGARGPQGVMGPPGKNGBAGFNKGERGPG 598

YQ 241 GPPGAVGPAKDGEAGAEGPPGPPGAPAGERGEEGPGSPGEGPLPGPAGPGGAEKPGEE 300

DB 599 GPPGTTPGAKNGDVLPGLPGLPAGDRGEPFSGSGFLQLGPPGPPAGENGKGP 658

YQ 301 GYPGDLGAPSPGARGEPPFPGERGVEGPPGPPGADGAPDDGAKGDAGAPGCGSE 360

DB 659 GPKGDIIGGPGFPKXGNGIPGERGFPQGPPTGARCGPFGAGSEGAKGPPPGAPG 718

YQ 361 GAPCLEGMPGERGAAGLPCKPKDRDAGPKADSGKDVRLGTGTRTPGPPGAPAGDX 420

DB 719 GLPGLQMFGERGASGSPGFKDGKGEKGDADGLFGARGERNVPLGPPGPPGPDGX 778

YQ 421 GESGPSGAPGTGARGAPDGRGEPGPPGAPGADGEPFGAKGEPGDAKGDAGDP 480

DB 779 GETGPAGAPGASRGPGERGEQGLPGLPAGFPAGPQNGEPGKGERGPPGLRGEAG 838

YQ 481 GPAGPAGPPGPTGDCVAPCAKAGRSAGPPGATGTFPGAAGRVGPPGSGD -----A 531

DB 839 GAAGPQGGFCAFPFPQGVKGRSGPGGPAAGFGARGPPGPPGNNNGKSGSGPPGVP 898

Query Match	Best Local Similarity	Score	DB 1	Length
Query	1	42.6%	Score 1921.5	DB 1: Length 1496
Db	396	64.0%	Pred. No. 2.6e-71	Mismatches 40; Indels 3; Gaps 1
Query	61	62.6%	Score 1921.5	DB 1: Length 1496
Db	453	62.6%	Score 1921.5	DB 1: Length 1496
Query	121	62.6%	Score 1921.5	DB 1: Length 1496
Db	513	62.6%	Score 1921.5	DB 1: Length 1496
Query	181	62.6%	Score 1921.5	DB 1: Length 1496
Db	573	62.6%	Score 1921.5	DB 1: Length 1496
Query	241	62.6%	Score 1921.5	DB 1: Length 1496
Db	633	62.6%	Score 1921.5	DB 1: Length 1496
Query	301	62.6%	Score 1921.5	DB 1: Length 1496
Db	693	62.6%	Score 1921.5	DB 1: Length 1496
Query	361	62.6%	Score 1921.5	DB 1: Length 1496
Db	753	62.6%	Score 1921.5	DB 1: Length 1496
Query	421	62.6%	Score 1921.5	DB 1: Length 1496
Db	813	62.6%	Score 1921.5	DB 1: Length 1496
Query	481	62.6%	Score 1921.5	DB 1: Length 1496
Db	873	62.6%	Score 1921.5	DB 1: Length 1496
Query	541	62.6%	Score 1921.5	DB 1: Length 1496
Db	933	62.6%	Score 1921.5	DB 1: Length 1496

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; X59251; CAA41205.1; -;
CC EMBL; BC007158; AAH07158.1; -;
CC EMBL; BC042503; AAH42503.2; -;
CC EMBL; K01832; AAA37331.1; -;
CC PIR; A43291; A43291;
CC MGD; MGI:88468; Colla2
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC ProDom; PD000007; C1g_helix; 6.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM0038; COUFI; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
CC SIGNAL 1 22 POTENTIAL.
CC PROPEP 23 85 AMINO-TERMINAL PROPEPTIDE
CC (BY SIMILARITY).
CC CHAIN 86 1108 COLLAGEN ALPHA 2(I) CHAIN.
CC PROPEP 1109 1372 CARBOXYL-TERMINAL PROPEPTIDE
CC (BY SIMILARITY).
CC MOD_RES 86 86 PYRROLIDONE CARBOXYLIC ACID (BY
CC MOD_RES 90 90 SIMILARITY).
CC INVOLVED IN CROSS-LINKING
CC CONVERTED TO AN ALDEHYDE GROUP THAT IS
CC (BY SIMILARITY).
CC N-LINKED (GLNAC..) (POTENTIAL).
CC CONFLICT 15 15 V -> A (IN REF. 4).
CC CONFLICT 1167 1167 R -> T (IN REF. 1).
CC SEQUENCE 1372 AA; 129557 MW; 0D17DF5D6C1452D1 CRC64;

Query Match 62.48; Score 1915; DB 1; Length 1372;
Best Local Similarity 64.28; Pred No. 4.5e-71;
Matches 349; Conservative 45; Mismatches 150; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPGAGAGPAGDAGDEFGAKGADGAGIAGAPGPGARGSPGE 60
DB 280 GPAGPAGRGVGLGLSGPVGPPGPNPTNGLTCAKATGLPGVAGAPGLGPRGIPGPA 339
QY 61 GPGGPPGKDSCEPPGAPGSKDGTGAKGEPGVGVEGPPGAPGEGKPGARGEPGLP 120
DB 340 GAAGATGARGLVGEPGAGSGENKGEPSVAGQPPGPGSGEGKRGSPGAGSAGPA 399
QY 121 GPPGERGPGSRGPPGADGVAGPKPGAGRGSPGAPGKSPGEGAGRPAGLPGAKLT 180
DB 400 GPPGLRGSPGSRGLPGADGRAGVMPGPNRGSTGTPAGIRPNFGDAGRPGEPGLMGRGLP 459
QY 181 GSPSGPDGKTGPPGAGEDRGPPGPPGARGVAGVFPKGNAGEPGKAGRGVVP 240
DB 460 GSPGNVFGSGKEGVPVLGIDRGPGTIGPAGPRGEAGNIFFPGPKGSPGPGKRGHP 519
QY 241 GPPGAVGPGAKDGSAGAGGPPGPPGAGRGEGEPGAPGSPGEGLPAGPFGAGPGE 300
DB 520 GLAGARGAPGPDGNNAGGPPGPPGVQGGKEQGPAGPPGFQGLPGSGTTGVEGKPGER 579
QY 301 GVPDGLGAPSGARGPPGPPGREGVGGPPGPPGADGAPDGDGAKGADGAPGAPGSE 360
DB 580 GLPGEFGLPGPAGPRGERTGSGAAGPSGPIGSRGTPGAPGPDGPNKGAGAVGAPGSA 639
QY 361 GAPLEGMPGPRGAAGLPGPKGDRGDAGPKGADGSPGKGVRLGTGPIGPPGAPGAPGDK 420
DB 640 GASGFGPLPGRGAGAGTGGKGEKGTGLRDTGNTGRTDGRGIPGAVGAPGAGSGDR 699
QY 421 GESGSPGAPGTGARGAPGDRGEPGPPGPPGAPGAGFAGPAGDGPAGKGEPPDAGAKGAGPP 480

DB 700 GEAGARPSGAPRGSGRGEVGPAGNFGAGACQPAKGEKGTGKFGENGIV 759
QY 481 GPAGPAGPPGPIGVGAPGAGKARGSGAPPGATGFGPAGAGRVGPPGSGDAGPPGPPGPA 540
DB 760 GPTGSGVAGPSGPPGPPGVPVSGKGDGPPGMTGFGPAGAGRTGPPGSGIAGPPGPGAA 819
QY 541 GKEG 544
DB 820 GKEG 823
DB
RESULT 14
CAL3_BOVIN STANDARD; PRT; 1049 AA.
ID CAL3_BOVIN
AC P04258;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain.
GN COL3A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-242.
RX MEDLINE=80026026; PubMed=488906;
RA Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
Kuhn K.;
RT "The covalent structure of calf skin type III collagen. I. The amino
acid sequence of the amino terminal region of the alpha 1(III) chain
(positions 1-222).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN [2]
RP SEQUENCE OF 243-422.
RX MEDLINE=80026027; PubMed=488907;
RA Dewes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. II. The amino
acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
(positions 223-402).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN [3]
RP SEQUENCE OF 423-571.
RX MEDLINE=80026028; PubMed=488908;
RA Bentz H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. III. The
amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
(positions 403-551).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN [4]
RP SEQUENCE OF 572-808.
RX MEDLINE=80026029; PubMed=488909;
RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. IV. The amino
acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
(positions 552-788).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN [5]
RP SEQUENCE OF 809-947.
RX MEDLINE=80026030; PubMed=488910;
RA Dewes H., Fietzek P.P., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. V. The amino
acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
(position 789-927).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN [6]
RP SEQUENCE OF 948-1049.
RX MEDLINE=80026031; PubMed=488911;
RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. VI. The amino
acid sequence of the carboxyterminal cyanogen bromide peptide alpha

RT 1(III)CB9B (positions 928-1028).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
 CC along with type I collagen.
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
 CC linked to each other by interchain disulfide bonds. Trimers are
 CC also cross-linked via hydroxyllysines.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC PIR; A02862; CGB078.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001007; VWF C.
 DR Pfam; PF01391; Collagen; 17.
 DR ProDom; PD000007; C1g_helix; 3.
 DR PROSITE; PS01208; VWF C 1; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.
 FT DOMAIN 1041 1049 NONHELICAL REGION (C-TERMINAL).
 FT MOD_RES 95 95 HYDROXYLATION.
 FT MOD_RES 107 107 HYDROXYLATION.
 FT MOD_RES 119 119 HYDROXYLATION.
 FT MOD_RES 938 938 HYDROXYLATION.
 FT MOD_RES 950 950 HYDROXYLATION.
 FT CARBOHYD 107 107 O-LINKED (GAL. . .).
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).
 FT DISULFID 1040 1040 INTERCHAIN.
 FT DISULFID 1041 1041 INTERCHAIN.
 SQ SEQUENCE 1049 AA; 93651 MW; 85EC33D1C66EC9A3 CRC64;
 Query Match 62.1%; Score 1905.5; DB 1; Length 1049;
 Best Local Similarity 62.7%; Pred. No. 9.1e-71;
 Matches 347; Conservative 36; Mismatches 161; Indels 9; Gaps 1;
 QY 1 GSEGEVGRGPGPGAGAGPAGDPCGNDGPGKAGDAGPGIAGAGFGFCARGPSPGE 60
 DB 204 GSSGAPGQGGEPGQHGAGPFPDPSGSGPGKGEWGPAGIFGAPGLIGARGPFGPP 263
 QY 61 GPGGPPGKGDSEGPCAPGSKGDTCAKGEPPGVGVEGPGPAGBEGKPGARGEPGPTGLP 120
 DB 264 GTNGVPGQGAAGEPGKNGAKGDPGRGREGAGSPGIAGPKGEDGKDGSPGPGANGLP 323
 QY 121 GPPRGGRGSGRFGPGADGAVGKPGAGRGSPGAGKSGPGEAGRGEGAGLPKAGLT 180
 DB 324 GAAGERGVFPGRGAGANGLPGEKGPDPGRGCGFGPAGPRGVAEPGRNGLPGGPGLRGP 383
 QY 181 GSPGSPGPKTGPAGPAGEDGRPPGPPGARGAGVGMFPKGAAGEPGKAGRGVP 240
 DB 384 GSPGPGSGNGKPGPGSQGTGRPPGSPGPRGQVGMFPKGNDAFGKNGRGGP 443
 QY 241 GPPGAVGPKDGEAGAGPAGPAGPAGRGEGPAGSGPFGFLPGPAGPGEAGKPGEE 300
 DB 444 GGPFGQPGAGKNGETGPQPPGPTGSGDKGTGPPGQGLPGLTSGPPGKNGKPGEP 503
 QY 301 GVPDGLGAPSGARGEPPFGRGVGPPGPPGADGADGADGADGADGADGADGADGADG 360
 DB 504 GPKGEAGAGTIGPKGSDGAPGAGRGPPGAGGPPGPRGAGPPGPGGKGAAGPFGPSGA 563
 QY 361 GAPLEGMPGERGAAGLPKGDGDPAGPKGADGSGPKDGVRLGTGPIGPPGAGAGDK 420
 DB 564 GTFGLQGMFGRGPGPGPKGDKGEPGSSGVDPAGPKDGRPGTGTGTPGPGAGQPGDK 623
 QY 421 GSGPFGPAGPTGARGAPDRGEPGPPGAGPAGPGADGEPGAGKGEFGDAGKADGPP 480
 DB 624 GSGAPGVPGIAGPRGPGRGEGEQGPPGPPGAGFPAGPQNGEPGKAGRGAPGKGGGPP 683
 QY 481 GPAGPAGPPIIDVGAAGKAGSAGSPGATGTFPGAAGRVGPPGPGSDAGPP----- 534
 DB 684 GAAGPAGGSGPAGPPGPGQGVKGERGSGPGGAGFGGPGGPGGSGNPNPFGSSGAP 743
 QY 535 ----GPPGPAKNG 544

744 GKDGPGPGPSNG 756

RESULT 15

DB Db
 ID CA13 HUMAN STANDARD; PRT; 1466 AA.
 AC P02461; O15112;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Proctor D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 RN [2]
 RP SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RN [3]
 RP SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RN [4]
 RP REVISIONS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RN [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RN [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CBS from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN [8]
 RP REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9]

RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RL procollagen";
RL Nucleic Acids Res. 12:9383-9394 (1984).
RN [10]
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha
RT 1(III)-C99 from type III collagen of human liver";
RL Biochemistry 20:2621-2627 (1981).
RN [11]
RP SEQUENCE OF 1176-1466 FROM N.A.
RX MEDLINE=85157600; PubMed=2579949;
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT (III) collagen. Partial characterization of the 3' end region of the
RT gene";
RL J. Biol. Chem. 260:4357-4363 (1985).
RN [12]
RP SEQUENCE OF 1161-1200 FROM N.A.
RX MEDLINE=86187804; PubMed=3754462;
RA Miskulin M., Dalgleish R., Kluge-Beckerman B., Renard S.I.,
RA Tolstoshev P., Brantly M., Crystal R.G.;
RT "Human type III collagen gene expression is coordinately modulated
RT with the type I collagen genes during fibroblast growth";
RL Biochemistry 25:1408-1413 (1986).
RN [13]
RP SEQUENCE OF 1-170 FROM N.A.
RT TSUVE=Placenta;
RX MEDLINE=88303360; PubMed=3405773;
RA Toman D., Ricca G., de Crombrughe B.;
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT of human proalpha 1(III) collagen";
RL Nucleic Acids Res. 16:7201-7201 (1988).
RN [14]
RP SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=89378752; PubMed=2777083;
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT "Cloning and analysis of the 5' portion of the human type-III
RT procollagen gene (COL3A1)";
RL Gene 78:255-265 (1989).
RN [15]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels";
RL Hum. Mutat. 9:300-315 (1997).
RN [16]
RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX MEDLINE=93293988; PubMed=8514866;
RA Tromp G., Wu Y., Prockop D.J., Madhaderi S.L., Kleinert C.,
RA Early J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
RA Cole C.W., Jaakkola P., Rymaenen M., Pearce W.H., Yao J.S.T.,
RA Majamaa K., Smulens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT in the triple-helical domain of type III procollagen are an
RT infrequent cause of aortic aneurysms";
RL J. Clin. Invest. 91:2539-2545 (1993).
RN [17]
RP VARIANT THR-698.
RX MEDLINE=91045136; PubMed=2235526;
RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA Wu Y., Ganguly A., Prockop D.J.;
RT "G to A polymorphism in exon 31 of the COL3A1 gene";
RL Nucleic Acids Res. 18:6180-6180 (1990).
RN [18]
RP VARIANT AORTIC ANEURYSM ARG-786.
RX MEDLINE=91056145; PubMed=2243125;
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT "A mutation in the gene for type III procollagen (COL3A1) in a family
RT with aortic aneurysms";
RL J. Clin. Invest. 86:1465-1473 (1990).
RN [19]
RP VARIANT EDS-IV ARG-828.
RX MEDLINE=94016385; PubMed=8411057;
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
RT "The substitution of glycine 661 by arginine in type III collagen
RT produces mutant molecules with different thermal stabilities and
RT causes Ehlers-Danlos syndrome type IV";
RL J. Med. Genet. 30:690-693 (1993).
RN [20]
RP VARIANT EDS-IV SER-957.
RX MEDLINE=89103135; PubMed=2492273;
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
RT "A single base mutation that substitutes serine for glycine 790 of
RT the alpha 1 (III) chain of type III procollagen exposes an arginine
RT and causes Ehlers-Danlos syndrome IV";
RL J. Biol. Chem. 264:1349-1352 (1989).
RN [21]
RP VARIANT EDS-IV VAL-960.
RX MEDLINE=95268429; PubMed=7749417;
RA Tromp G., de Paeppe A., Nuytinck L., Madhaderi S.L., Kuivaniemi H.;
RT "Substitution of valine for glycine 793 in type III procollagen in
RT Ehlers-Danlos syndrome type IV";
RL Hum. Mutat. 5:179-181 (1995).
RN [22]
RP VARIANT EDS-IV GLU-1014.
RX MEDLINE=92316511; PubMed=1352273;
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA Pope F.M.;
RT "A single base mutation in the gene for type III collagen (COL3A1)
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT syndrome type IV. An unaffected family member is mosaic for the
RT mutation";
RL Hum. Genet. 89:414-418 (1992).
RN [23]
RP VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT "Single base mutation in the type III procollagen gene that converts
RT the codon for glycine 883 to aspartate in a mild variant of
RT Ehlers-Danlos syndrome IV";
RL J. Biol. Chem. 264:19313-19317 (1989).
RN [24]
RP VARIANT EDS-IV VAL-1077.
RX MEDLINE=91374480; PubMed=1895316;
RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
RA Pope F.M.;
RT "Characterisation of a glycine to valine substitution at amino acid
RT position 910 of the triple helical region of type III collagen in a
RT patient with Ehlers-Danlos syndrome type IV";
RL J. Med. Genet. 28:458-463 (1991).
RN [25]
RP VARIANT EDS-IV GLU-1173.
RX MEDLINE=93022543; PubMed=1357232;
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
RT "Query Match 61.9%; Score 1901; DB 1; Length 1466;
RT Best Local Similarity 63.1%; Pred. No. 1.7e-70;
RT Matches 345; Conservative 37; Mismatches 159; Indels 6; Gaps 2;
RT 1 GSEGEVGRGPGPGPAGAGDPADGDEPAKAGDAGPAGAGPFGARGPSGPE 60
RT 360 GSNAGPGQGRGPGPGAGAGQPGPGPGINGSPGKGEMPGAGIPGAPGLMGARGPGPA 419
RT 61 GRGGPPGKSGDGGPGAPGSGKGTGAKGPGVGVGPPGPGAGGEGKPGARGEPGTGLP 120
RT 420 GAGAPGLRGAGGPGKNGKAGGPGRGGEAGIPGVPGAKGDKGDSFGFGANGLP 479

